

-1499-

```

      MEL Q KRG+F+ M Y+D NEV++IY +PLAEIV++PFD+LKSST+GYASFDY++ Y+
  5  Sbjct: 429 MELCGKRGNFIDMQYLDANKVSIYYDMPLAEIVTEFPDQLKSSSTKGYSFDYELIGYKP 488

  Query: 488 SOLVKMDILLANGDKVDALSPFVHKPEFAYERGKIIVBKLLKIIIPRQQFVEPTQAAIQGKIV 547
      S+LVKMDI+IANG+K+DALSGFVH++AYERGK+IVBKLLK++IPRQQFVEPTQAAIQGKIV
  5  Sbjct: 489 SKLVKMDIMLNGEKIDALSPFVHRYAYERGKIVVEKLKELIPRQQFVEPTQAAIQGKIV 548

  Query: 548 ARSDIKALKRNVLAKCYGGDVSRRKKILEKQKAGKKRMKAIGSVVEVPQEAFLSVLSMDDO 607
      ARS IKA+RKNVLAKCYGGD+SRKKILEKQK GK+RMK +GSVEVPQEAFL+VVL MDD
  10 Sbjct: 549 ARSTIKAMRKNVLAKCYGGDISRRKKILEKQKAGKKRMKQVGSVEVPQEAFLVLMDDO 608

  Query: 608 TKK 610
      KK
  15 Sbjct: 609 PKK 611

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 587/610 (96%), Positives = 601/610 (98%)
  20 Query: 1 MNIEDLKROEKIRNFSTIAIHDSGKSTLADRILEKTETVSSREMQAQLLDSMDLERERG 60
      MN +DLKROEKIRNFSTIAIHDSGKSTLADRILEKTETVSSREMQAQLLDSMDLERERG
      Sbjct: 1 MNSQDLKKRQEKIRNFSTIAIHDSGKSTLADRILEKTETVSSREMQAQLLDSMDLERERG 60

  Query: 61 ITIKNAIEINLYTAKDGETYI FHLIDTPGHVDFTEYVSRSLAACBGAILVDDAAQGIEAQ 120
      ITIKNAIEINLYTAKDGETYI FHLIDTPGHVDFTEYVSRSLAACBGAILVDDAAQGIEAQ
  25 Sbjct: 61 ITIKNAIEINLYTAKDGETYI FHLIDTPGHVDFTEYVSRSLAACBGAILVDDAAQGIEAQ 120

  Query: 121 TLANYVALINDLEILFPVINKIDLPADPERVRAEVEDVIGLDASEAVLASAKAGIGIEE 180
      TLANYVALINDLEILFPVINKIDLPADPERVR EVEDVIGLDASEAVLASAKAGIGIEE
  30 Sbjct: 121 TLANYVALINDLEILFPVINKIDLPADPERVRHEVEDVIGLDASEAVLASAKAGIGIEE 180

  Query: 181 ILEQIVEKVPAPTGEVDAPLQALIPDSVYDAYRGVILQVRIVNGMVGQDKIQMMSNGKT 240
      ILEQIVEKVPAPTGV+VDAFLQALIPDSVYDAYRGVILQVRIVNG+VKGDKIQMMSNGKT
  35 Sbjct: 181 ILEQIVEKVPAPTGEVDAPLQALIPDSVYDAYRGVILQVRIVNGIVKQDKIQMMSNGKT 240

  Query: 241 FDTVTEVGIFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDTITLANNPAIEFLHGYKQM 300
      FDTVTEVGIFTPKAVGRDFLATGDVG+AAIKTVADTRVGDT+ITLANNP A E L HGYKQM
  40 Sbjct: 241 FDTVTEVGIFTPKAVGRDFLATGDVGVAASIKTVADTRVGDTITLANNPAKALEHGYKQM 300

  Query: 301 NPMVFAGLYPIESNKYNDLREALEKLQINDASLQFEPETSQALGFGRFCGLGLLHMVDI 360
      NPMVFAG+YPIESNKYNDLREALEKLQINDASLQFEPETSQALGFGRFCGLGLLHMVDI
  45 Sbjct: 301 NPMVFAGLYPIESNKYNDLREALEKLQINDASLQFEPETSQALGFGRFCGLGLLHMVDI 360

  Query: 361 QERLEREFNIDLINTAPSVVYHVMTDGEKLEVSNPSEFFDPDTRVDSIEEPVKAQIMVP 420
      QERLEREFNIDLINTAPSVVYHV+TTD +M+EVSNPSEFFDPDTRV I EEPVKAQIMVP
  50 Sbjct: 361 QERLEREFNIDLINTAPSVVYHVHVTDDMEIVSNPSEFFDPDTRVPI EEPVKAQIMVP 420

  Query: 421 QEPUGAVMELAQKRGDFVTNDYIDONRVNIYQIPLAEIVDFDFDLKLSSTRGYASFDY 480
      QEPUGAVMEL+QKRGDFVTNDYIDONRVNIYQIPLAEIVDFDFDLKLSSTRGYASFDY
  55 Sbjct: 421 QEPUGAVMELSQKRGDFVTNDYIDONRVNIYQIPLAEIVDFDFDLKLSSTRGYASFDY 480

  Query: 481 RISEYRRSQLKMDILLANGDKVDALSPFVHKPEFAYERGKIIVBKLLKIIIPRQQFVEPTQAA 540
      ++SEYRRSQL KMDILLANGDKVDALSPFVHKPEFAYERGK+IV+KLLKIIIPRQQFVEPTQAA
  60 Sbjct: 481 DMSEYRRSQLKMDILLANGDKVDALSPFVHKPEFAYERGKIIVBKLLKIIIPRQQFVEPTQAA 540

  Query: 541 AIGQKIVARSIDIKALKRNVLAKCYGGDVSRRKKILEKQKAGKKRMKAIGSVVEVPQEAFLS 600
      AIGQKIVARSIDIKALKRNVLAKCYGGDVSRRKKILEKQKAGKKRMKAIGSVVEVPQEAFLS
  65 Sbjct: 541 AIGQKIVARSIDIKALKRNVLAKCYGGDVSRRKKILEKQKAGKKRMKAIGSVVEVPQEAFLS 600

  Query: 601 VLSMDDDDKK 610
      VLSMDDO KK
  70 Sbjct: 601 VLSMDDDTYKK 610

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1500-

**Example 1370**

A DNA sequence (GBSx1455) was identified in *S.agalactiae* <SEQ ID 4189> which encodes the amino acid sequence <SEQ ID 4190>. This protein is predicted to be awd gene product (ndk). Analysis of this protein sequence reveals the following:

```

5   Possible site: 42
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.2097 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAF57188 GB:AE003779 awd gene product [Drosophila melanogaster]
    Identities = 73/136 (53%), Positives = 100/136 (72%), Gaps = 5/136 (3%)

    Query: 2   EQTFPMIKPDGVKRGFIGEVISRIERRGFSIDRLVRYACADILKRHYAELTDRPFPPFL 61
              E+TF N+XPDGV+RG +G++I R E++GF + L+ +A ++L++HYA+L+ RPPFP L
    Sbjct: 25  ERTFIMVXPDGVQRGLVQKILIERFBQKGFILVAKFTWASKELLEKHVADLSARPPFGL 84

20  Query: 62   VDMYSGPVLIGVISGEEVISTWRTMMGSTNPKDALPGTIRQDFAQSPNQATCNIVHG 121
              V+YN SGPV+ V G V+ T R M+G+TNP D++PGTIRGDF Q NI+HG
    Sbjct: 85  VYNNYSGPVMVMWVGLVVKITGRQMLGATNPADSLPGTIRGDFC-----IQVGRNIING 139

25  Query: 122  SDSPEASATREIAWPN 137
              SD+ ESA +EIA+WPN
    Sbjct: 140 SDAVESASKEIALWPN 155

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4191> which encodes the amino acid sequence <SEQ ID 4192>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 22
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2913 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40  Identities = 30/48 (62%), Positives = 35/48 (72%)

    Query: 87   MMGSTNPKDALPGTIRQDFAQAPSPNQATCNIVHGSDSPEASATREIAI 134
              MM TNPKDAL GTIR +FAQAP + N+VHGS S +SA REIA+
    Sbjct: 1   MMRVYNPKDALOGTIRENFAQAPGDDGIFNMVHGSHSRDSARREIAL 48

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1371**

A DNA sequence (GBSx1456) was identified in *S.agalactiae* <SEQ ID 4193> which encodes the amino acid sequence <SEQ ID 4194>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 15
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
55  bacterial cytoplasm --- Certainty=0.2734 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-1501-

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4195> which encodes the amino acid sequence <SEQ ID 4196>. Analysis of this protein sequence reveals the following:

Possible site: 16

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1985(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 28 SFGTIRNSTALKQLTLDLSLNLLSFGTIRNSTALK 61

SFGTI+NS ALKQ + +N SFGTI+NS ALK

Sbjct: 7 SFGTIQNSIALKQKQAEINQRSFGTIQNSIALK 40

20 Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 6 SFGTIRNSTALKLYAKQSPAFRSFGTIRNSTALK 39

SFGTI+NS ALK A++ RSFGTI+NS ALK

Sbjct: 7 SFGTIQNSIALKQKQAEINQRSFGTIQNSIALK 40

25

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1372

A DNA sequence (GBSx1457) was identified in *S.agalactiae* <SEQ ID 4197> which encodes the amino acid sequence <SEQ ID 4198>. Analysis of this protein sequence reveals the following:

Possible site: 16

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.1407(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4199> which encodes the amino acid sequence <SEQ ID 4200>. Analysis of this protein sequence reveals the following:

Possible site: 37

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 154/221 (69%), Positives = 187/221 (83%)

Query: 1 MIKINFPIIDPLNLNATLTIEDVSYSLLVGHFYCYQYDVEHLKLPDQKSLKATEL 60

++ +NF +LDEP+ L TIL +EDV V+S +V+ YQY+ D LK FD K K+K +E+

-1502-

Sbjct: 8 LNNLNFSLLDPEIPLRGTTILVLEEDVCFVSKIVQYCYQYREDSLEKFFDHRMKTIKESKI 67  
 Query: 61 MLVTDILGYDVNSAPILKILKINGLENQFNEKFEVKSMVKKLAANTITELIAPFCLNELDL 120  
 MLVTDILG+DVNS+ ILKLIH DLE+QFNEKFEVKSM++KL ATITELI FECLNELDL  
 Sbjct: 68 MLVTDILGFDVNSSTILKLIHAULESQFNEKFEVKSMIDKLVATITELIVFECLNELDL 127  
 Query: 121 EYDEIKILELIALGVKIETQSDTIFEKCFEIIQVYHVI/TKNNLVPVNSGAYLTKDEVI 180  
 EYDEI IDELIK+LGK+STQSDTIFEK RI+Q++ YLTCK LL+FVNSGA+LTKDEV  
 Sbjct: 128 EYDITITILEIKSLGKVKETQSDTIFEKCLEIQL FKYLTKKKLLIPVNSGAFLYTKEVA 187  
 Query: 181 KLCEYINLMQKSVLFLEPRKLYDLQYVYIDKDYFLIGHNMV 221  
 L EYI+L +VLFLEPR LYD PQY++D-DYFLI +NMV  
 Sbjct: 188 SLQYETISLNLVTLFLEPRSLYDFPQYILOEDYFLITKNMV 228  
 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1373

A DNA sequence (GBSx1458) was identified in *S. agalactiae* <SEQ ID 4201> which encodes the amino acid sequence <SEQ ID 4202>. Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0842 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9783> which encodes amino acid sequence <SEQ ID 9784> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB83918 GB:AL162753 hypothetical protein NMA0629 [Neisseria meningitidis Z2491]  
 Identities = 45/104 (43%), Positives = 65/104 (62%), Gaps = 2/104 (1%)  
 Query: 4 RYMRMILMFDMPETAEERKAYRKFRKFLLSBGFIMHQFSVYSKILLANTANNAMIGRLK 63  
 ++NR+I+ FD+P TA +REA +FR+FL +G+ M Q SVYS+++ + RL  
 Sbjct: 5 KFRIRIYFFDLPTVITAAKRAANQFRQFLLDGYQMLQLSVYSRIVKGRDSLQKHENRLC 64  
 Query: 64 VNNPKKGNITLTLVTEKQFARMVYLHGERNT--SVANSDSRLVF 105  
 N P++G+I L +TEKQ+A M L GR T NSD L+P  
 Sbjct: 65 ANLQCGSIRCLTEITKQYAMKLLLGELATQKKKVNSSQLLLF 108

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4203> which encodes the amino acid sequence <SEQ ID 4204>. Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0822 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 97/112 (86%), Positives = 107/112 (94%)  
 Query: 1 MSYRYNRMILMFDMPETAEERKAYRKFRKFLLSBGFIMHQFSVYSKILLANTANNAMIG 60  
 MSYRYNRMILMFDMPETAEERKAYRKFRKFLLSBGFIMHQFS+YSKILLANTANNAMIG



-1503-

Subjct: 1 MSYRIMRMILMFDMPDITASSRKAYRKFRKFLLESGFIMHQFSIYKLLLNNTANNAMIG 60

Query: 61 RLKVNPKKGNITLLFVTEKQFARMVYLGERNTSVANSDSRLVFLGDSYDQ 112

5 Subjct: 61 RLREHNPKNKGNITLLFVTEKQFARMVYLGERNNTANSRDLRVFLGSAFDE 112

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1374

- 10 A DNA sequence (GBSx1459) was identified in *S.agalactiae* <SEQ ID 4205> which encodes the amino acid sequence <SEQ ID 4206>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3185 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB83919 GB:AL162753 hypothetical protein NMA0630 [Neisseria meningitidis Z2491]

Identities = 71/224 (31%), Positives = 122/224 (53%)

- 25 Query: 4 WRTVVVNTSKLSYKNNHLIPKDSYQTMHILSEIDILIMETTDIVLSTMLIKRLVDENI 63  
                   WR++++ KLS + L+ + + + + L + I + + I + E + + + + L+ L +  
 Subjct: 3 WRSLLIQNGSKLSLQRQLLIQNGESHVPLEDAVII IENRSTLI TAPLLSALAEHGA 62
- 30 Query: 64 LVIFCDKRLPTAMLMPPYARHDSLSQLSRQMSWIEDVKADVWTSIIAQKILNQSPLYGE 123  
                   ++ CD++ LP ++ PY H L Q++ E + K + W I+ QKILNQ+ F E  
 Subjct: 63 TLLTCDEQELFCQGWLPYQYHRQLIKLKLQNLISEPLKQLWQHIVRQKILNQAFVADE 122
- Query: 124 CSFFPKSQSIMNLYHDLSEFPDPSNREGHAAITYPNTLFGNDFSRQDNFINAGLDYGYSL 183  
                   ++ + L ++ D NRE AA + YF LFG F+R +N +NA I+Y Y++
- 35 Subjct: 123 TGNDLAAKRLRTLASEVRSGETGNREQAQAALYFQALGFEKPTRDNNAVALNITYAV 182
- Query: 184 LLSMFAREVVKQGCMTQPGKHKANQFNQNLASDIMEFPFRPIVD 227  
                   L = AR + G + GL H ++ N FNLA D +EP RP+ D
- 40 Subjct: 183 LRAAVARALTYGNLPAALGFHRSLEMPNLADDFIEPLRPLAD 226

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4207> which encodes the amino acid sequence <SEQ ID 4208>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3185 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 50 An alignment of the GAS and GBS proteins is shown below.

Identities = 239/289 (82%), Positives = 271/289 (93%)

- 55 Query: 1 MAGWRTVVVNTSKLSYKNNHLIPKDSYQTMHILSEIDILIMETTDIVLSTMLIKRLVD 60  
                   MAGWRTVVVNTSKLSYKNNHLIPKDSYQTMHILSEIDILIMETTDIVLSTMLIKRLVD 60  
 Subjct: 1 MAGWRTVVVNTSKLSYKNNHLIPKDSYQTMHILSEIDILIMETTDIVLSTMLIKRLVD 60
- Query: 61 ENILVIFCDKRLPTAMLMPPYARHDSLSQLSRQMSWIEDVKADVWTSIIAQKILNQSPLY 120  
                   EN+LVIFCDKRLPTAMLMPPY RHDSLSQL +QMSW E V+K+ VWI+IAQKILNQS Y

-1504-

Sbjct: 61 ENVLVIPCDKRLPTAMLPFYGRHDSLQLGKQMSSTFVKSQVWTTTIAQKILNCSQCY 120

Query: 121 LGRCSYFEKSSQSIMNLYHDLPEFDPNSNREGHAARIYFWTLPGNDPSRQDNPTNAGLDYOG 180  
 LG CS+FEKSSQSIM+LYH LE FDPNSNREGHAARIYFWTLPGNDPSR+ ++PTNAGLDYOG

5 Sbjct: 121 LGRCSYFEKSSQSIMDLYHGLNFDPNSNREGHAARIYFWTLPGNDPSRDLHPINAGLDYOG 180

Query: 181 YSLLSMFAREVVGCGMCTQFGLKHANQFNQFNLASDIMEPFPIVDRIIYENRQSDPVK 240  
 Y+LLSMFAREVVG CGMCTQFGLKHANQFNQFN ASDIMEFPF+VD+I+YENR P K

10 Sbjct: 181 YTLLSMFAREVVGCGMCTQFGLKHANQFNQFNASDIMEPFPIVDRIIYENRQSDPVK 240

Query: 241 MKRELPSMFSTSYNGKEMYLNIIVSDYTKKVK+LN++G+G+PBFRI 289  
 +KREL+PS+T+SYNGKEMYL+NI+SDYTKKVK+LN++G+G+PBFRI

Sbjct: 241 IKRELPTLSDTSYNGKEMYLNIIVSDYTKKVK+LN++G+G+PBFRI 289

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1375

A DNA sequence (GBSx1460) was identified in *S. agalactiae* <SEQ ID 4209> which encodes the amino acid sequence <SEQ ID 4210>. Analysis of this protein sequence reveals the following:

20 Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1109 (Affirmative) < succ>  
 bacterial membranes --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAB73943 GB:AL139078 hypothetical protein Cj1523c [Campylobacter  
 jejuni]  
 Identities = 165/746 (22%), Positives = 291/746 (38%), Gaps = 115/746 (15%)

Query: 318 LSASMIQRYDEHREDLQKLQFVQASLPEKTYQEI--FADSSKQYAGYIEGKTNQBAFYK 375  
 L+ S +R + L LK + Y++ F+S Y G + E ++

35 Sbjct: 50 LARSARKRLARRKARLNHLHLIANEFKLNYEDYQSPDESGLKAYKGLISP--YELRFR 107

Query: 376 YLSKLLTKQEDSENFLB--KLNEDFLRKQRTFDNNGSIPHQVHLTELKATIRQS----- 428  
 L+LL+EQ+ + L K + D ++ + G+I + E K + QS

40 Sbjct: 108 ALNELLSKQDFARVILHIAGRGYDDTKNSDDKKGAILKATKQNEER-LANTQSVGEYL 166

Query: 429 --EYTFPLEKQDRIKILTPRIPTY-----IGPLAREKSDPAM-MTRKTDSDI 474  
 EY+ KEN + + Y + + ++ +F + ++K ++ +

45 Sbjct: 167 YKEPTQFKENSKEPTNVRNKKGSYERCIASFLKDEHLKILFKQREPGSPSKKFEERV 226

Query: 475 RPNWNEDELVDKEKSAEAFIHRMINNDPYLEBEKVLPHSLIYEKPTVYNELTKV--RYKN 532  
 P +++ + FH + N P + EK PK+S + F + + +N

50 Sbjct: 227 LSVAFY-----KRALKDFSHLVGNCSFTT-DEKRAFKNSPLAFMFVALTRIINLNLNKN 280

Query: 533 HQQETYPFDSNIIQEIYFDGVFKSHKRVSK--KKLLDPLAKYEFERFIVDVGIDKHNKAP 590  
 +G Y D + + V K K KKL L+ +E R +

55 Sbjct: 281 TEGILYTKD--LWALLNEVLKNGTLTYKQTKLLG-LSDDT-----FKGEKGY 328

Query: 591 NASLGYTHLEKILDKPLDNPDSISLBDIVQTLLTFEDREMIKVRLENYVLDLFTSGL 650  
 Y + K L + L D L +I + +TL +D +K L Y +

Sbjct: 329 FIEFKKYKEPKALGSHNLQD----LNIAKDITLIKDEIKLKALAYD--LNQNGI 382

Query: 651 KKLRYRHYTGWRGLSAKLINGIRDK--ESQKTLVDLDDGRSNRNFQMLINDGLSPKS 708  
 L + + + S K + + E +K D+ + N IN+D P

60 Sbjct: 383 DLSKLEPKDHLNISFKALKLVTPLMLEKKK-----YDEACNRLNLKVAINEDKKDPLP 436

Query: 709 IISKAQAGSHSDNLKEVVGELAGSPAIKKGLQSLKTVDELVKVMGVEFQIVVENAREN 768  
 ++ N P + I + K++ L+K G + +I +E+ARE

-1505-

Sbjct: 437 AFNETYYKDEVTH-----PVVLRAIKEYRKVLNALLKKYG-KVHKINILAREV 484

Query: 769 QTNQGR---RNSRQRYKLLDDG---VKNLASDLNG-NILKEYPTDNQALQNERLFLY 820  
+ R + + YK D + L +N NILK L L+

5 Sbjct: 485 GGNHSQRKIKIKQENYKAKQDALECEKGLKINSKILK-----LRLPK 531

Query: 821 LQNGRDMYTGALDIDNLSQ---YDIDHIIPQAFIKDSDINRVLSSAKNRKKSDDVPS 877  
Q Y+GE + I +L +IDHI P + DDE N+VIV + +N+ K + P

10 Sbjct: 532 EQKEFCAYSGEKIKI SDLQDEKMLEIDHIYPIGRSPDQSYMNKVLVFTQMQEKLQNTF 590

Query: 878 LSIVKDKCKVFKKL--LDKLMGSRKYDNLTGAERGGLTSDDKARFIQRLVETRTQITKH 935  
E + W+K+ L L +++ L K ++ F R L +TR I +

Sbjct: 591 FFAFGNCSAQMKEVLAKNLPKKQRILDK----NYKDKBQGNFKDRNLNDITRYLRL 646

15 Query: 936 VARI-----LDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGYKIREVANY 986  
V L + N +L+ ++ KV + L S R +GF N+

Sbjct: 647 VLVNPKYDLDPLFLSDDENTKLNLT-QGSKSVVEAKSGMLTSLALNTWGFSAKDRNNHL 705

Query: 987 HHAIDAVILNAVVAAILTYKPLEFE 1012  
HHA DA + A +I+ + + E

20 Sbjct: 706 HHAIDAVIAYANNSIVKAFEDFKKE 731

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4211> which encodes the amino acid sequence <SEQ ID 4212>. Analysis of this protein sequence reveals the following:

25 Possible site: 61  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0973 (affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 881/1380 (63%), Positives = 1088/1380 (78%), Gaps = 22/1380 (1%)

35 Query: 1 MNKPYSIGLDIGTNSVGMSIITDVKVPKKNRVLGNTDKEVILKGLIGALLFDGGSTAA 60  
M+K YSIGLDIGTNSVGW++ITD+YKVP+KK +VLGNTD+ IKKGLIGALLFD G TA  
Sbjct: 1 MNKPYSIGLDIGTNSVGMAVITDEYKVPKPKFVLGNTDRHSIKKGLIGALLFDGGSTAE 60

40 Query: 61 DRRLKRTARRRYTRRRNRILYLOEIPAREMSKVDDSPFHRLEDSFLVEEDKRGSKYPIPA 120  
RLKRTARRRYTRR-NRI YLQEI P+ EM+KVDDSPFHRLE+GFLVEDK+ ++PIF  
Sbjct: 61 ATRLKRTARRRYTRKKNRI CYLQEIFSNEMAKVDDSPFHRLEESFLVEEDKKKHKHPIG 120

45 Query: 121 TLCEEDYHKEFSTIYHLRKLAKLKKKADLRILYIALAHIKFRGHFLIEDSPVVRNT 180  
+ +E YHKK+ TIYHLRK+L D +KADRLIY+ALAH+IKFRGHFLIE D + N+  
Sbjct: 121 NIVDEVAYHKEYPTIYHLRKLKLVDSDDKADRLIYIALAHMIKFRGHFLIED-LNFDHS 179

Query: 181 DISKQYQDFLEIPNTTFFNNOLLSQNVVAILTDKISKSAKKDRILAQYFNKQSTGIFA 240  
D+ K + ++ +N FE N + ++ + VD +AIL+ ++SKS + + ++AQ P +K G+ F  
50 Sbjct: 180 DVDKLPQLVQTVYNQLFEENPIKASGVDAKAILSARLSKSRRLNLIQLGPEKKKGLFEG 239

Query: 241 EFLKLIWQADFKKYFNLEDKTPLOPKADSYDEDLNLLQIQDFADFLSAAKKIYDS 300  
+ L +G +PK F+L + LQ +KD+YD+DL+NLQ QGD+ +ADLF PAK L D+  
55 Sbjct: 240 NLIALSIGLTPNPKSNFDLAEDAKLQSKRTYDIDDLNLLAQIQGDVADFLPAAKNLSDA 299

Query: 301 VLLSGILTVLDLTKAPLSASMIORYDEHREDLKOLKQFVKASLPEKYQEIFADSGKDY 360  
+LLS IL V TKAPLSASMI+RYDEX +DL LK V+ LPEKY+EIF D SK+GY  
Sbjct: 300 ILLSDILKVNTEITKAPLSASMIKRYDEHREDLTLKALVRQQLPEKYKEIFPDQSKNGY 359

60 Query: 361 AGYIEGKTNGEAFYKYLKSKLLTKQSDENFLEKIKNEEDFLRKQRTFDNGSIPQHVHTEL 420  
AGYI+G +QE FYK++ +L K + +E L K+ ED LRKQRTFDNGSIPQ+HL EL  
Sbjct: 360 AGYIDGASQEEFYKFKIPLEKMDGTRELLVKNREDLLRKQRTFDNGSIPQHVHTEL 419

Query: 421 KAILRQSEYYPFLKENODRIEKLITFRIPYYIGPLAREKSDFAWNRKTRTDDSIKRPINFE 480  
AI+REQ ++YPLK+N+++IEKLITFRIPYY+GPLAR S FAWNTRK+++I PWNFE

65

-1506-

Sbjct: 420 HAILRRQSDFYFPLKDNREKIEKILTFRIPTYVGLPARGNSRFAMTRKSEPTITWNPFE 479

Query: 481 DLVDKRSASAPFIHMTNNDPFLPKVILPKHSILYEKPTVYNELTKVRVGN-QGRYV 539  
++VDK SA++FI RMN D LP EKVLPHSIL+YE FTVYNELTKV+Y E + P

5 Sbjct: 480 EVDVKGASASQPIRMNTNFDKNLPEKVLPHSILYEKPTVYNELTKVKVYTEGMRKPAF 539

Query: 540 FDSNIKQEIFDGVGHEKRRVSKKLLDPLAKVEYSEFPIVDVIGLOKKNKAPNASIGCTYHD 599  
K+ I D +FK +RRV+ K+L + K+ E F V+ G+ + PNASIGCTYHD

10 Sbjct: 540 LSGEQKKAIVDLLFTKTRKVTYVQLKEDYFKKIECFDSVRI+SGVEDR---PNASIGCTYHD 596

Query: 600 LEKIL-DKDFLDMPDNISILEIDVOTLTFLFEDRMIKKRLNRYKOLFTPSQLKKIVRRHY 658  
L KI+ DKDFLDN -NS ILEIDV TLTFLFEDRMI++RL+ Y LF + +K+L RR Y

Sbjct: 597 LKIIKDKDFLDHNEEDILEIDIVATLTFLFEDRMIHERLKYTAHLFDKVMWQLKRRY 656

15 Query: 659 TGWGRLSAKLINGIRDKSSQKLTLDYLDIDGSSNRNFWQLINDDGLSPKSIISKAGQSH 718  
TGWGRLS KLINGIRDK+S KTIID+L DG +NRNFWQLI+DD L+FK I KQ

Sbjct: 657 TGWGRLSKLINGIRDKQSGKITLDFLKSDGFANRNFWQLIHDDSLTFEDIKQAGVSOQ 716

20 Query: 719 SDNLKEVVGELAGSPAIKGILQSLKIVDELKVMG+YEPQIVVBPAREMQTTNQRRN 777  
D+L E + LAGSPAIKGILQ+K+VDELKVMG ++PE IV+BPAREMQTT ++N

Sbjct: 717 GDSLHEHLANLAGSPAIKGILQTVKVVDELKVMGRHKEPVIVAREMQTTQKQJN 776

Query: 778 SRQRYKLLDGVKNLADLGNLKEYPTNQALQNERFLYTLQNGRMWYTGALDIDN 837  
SR+R K +++G+K L S ILKE+P +N LQNE+L+LYTLQNGRMWY +LDI+

25 Sbjct: 777 SRQMKRIEGBGKELGS----QTLKEHPVENTLQNEKLYTLQNGRMWYDQGLDNR 832

Query: 838 LQQDIDHIIPOAFIKDSDIKNRVULVSSAINRGKSDDVPSLKIIVDKCVFWKILLDAKLM 897  
LS YD+DHI+PQ+P+KDDSDIN+VL S KNRGKSD+PS E+VK K +N+L+L+AKL+

30 Sbjct: 833 LQDIDVDHIVPQSLKDDSDINKVLTRSDKNRGKSDNVPSEVVKMKRYNRQLNAKLI 892

Query: 898 SQRYKINLTAKERGSLSDDKARPTQROLVETROITKHVARILDERFNMLDQSKRRIK 957  
+QRK+DNLTKAERGSL+ DKA PT+RQRLVETROITKHVA+ILD R N + D + IR+

Sbjct: 893 SQRYKINLTAKERGSLSDDKAGFITKQRLVETROITKHVAQLDSDNMTKYDENDKLIRE 952

35 Query: 958 VKIVITLKSMLVSNFRKEGPFYKIREVNYHHADAYLNMAVAKALITYPQLESPFVYGD 1017  
VK++TLKS LVS+FRK+F FYK+RE-NYHHADAYLNMAV A++ KYP+LE SPFVYGD

Sbjct: 953 VKVITLKSGLVSDFRIDPFYKVRBNYHHADAYLNMAVGTALIKKYPKLESPFVYGD 1012

40 Query: 1018 YPKYN-----SYTKRKSATEKLFYFSNIMNFFKTKVLADGTVVVKDIEVNRDGEI 1070  
Y Y+ S + AT K FFSNIMNFFK+TLA-G + + IE N +D+TGEB

Sbjct: 1013 YKVIDVRMIKSEGEIGKATKAYFFYSNIMNFFKTEITLANGEIRKPLEIENKGTGEB 1072

Query: 1071 VNDKGGHPATVRKVLSPQVNIKKTEITQTGGFSKESILAHGNSDKLPRKTKDIYLDPK 1130  
VNDK + FATVRKVLSP PQ NIVKTE+QTGGFSKESIL NSDKL RK KD DPK

45 Sbjct: 1073 VNDKGRDPATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARK-KD--NDPK 1129

Query: 1131 KYGGFDSPTVAYSVLVADIKKGKAQKLETVTELLGITTIMERSFEKNPSAFLESKGVIN 1190  
KYGGFDSPT VAYSVLVA ++K+K+L+K+V ELLGITTIMERS FEKNP FLE+K+K+V

50 Sbjct: 1130 KYGGFDSPTVAYSVLVAKVKGKSKLKSVELLGITTIMERSFEKNPIDFLKARYKE 1189

Query: 1191 IRADKLIILPKYSLPELEWGRKRLASAGELQKGNKALPQTQMKFLYLASRVNSKXPK 1250  
++ D +I LPKYSLELEWGR+R+LASAGELQKGNKALP++++ PLYLAS Y + XG P

Sbjct: 1190 VKKDLIIKLPKYSLELEWGRKRLASAGELQKGNKALPQKVNFLYLASHYKELKGP 1249

55 Query: 1251 EEIEKKQSFVGHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENSVDLEANN 1310  
E+ E+KQ PV GH Y D+I++ I+FSKRVILADANL+K+ Y +++ +E A N

Sbjct: 1250 ENEQKQSFVGHVGHYDEIEQISEFSKRVILADANLKVLSAYNHRDK-PIREQAN 1308

60 Query: 1311 IINLFTFPLSGAPAFKPFDKIVDRKKYTSKEVLNSTLHQSTGLYETRIDLGLGSD 1370  
II+LFT T+LGAPAFK+PD +DRKRYTSKEVL++TLHQSTGLYETRIDL +LG D

Sbjct: 1309 ITHLFTLNLGAPAFKPYDITTDKRYTSKEVLDAITLHQSTGLYETRIDLGLGSD 1368

SEQ ID 4210 (GBS317) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 2; MW 179.3kDa) and in Figure 159 (lane 5 & 6; MW 180kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 3; MW 154.3kDa) and in Figure 159 (lane 9 & 10; MW 154kDa).

-1507-

GBS317-GST was purified as shown in Figure 224, lane 9-10. GBS317-His was purified as shown in Figure 222, lane 9.

GBS317N was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 2-4; MW 116kDa).

5 GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 6-8; MW 92kDa).

GBS317dN was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 7; MW 116kDa). Purified GBS317dN-GST is shown in Figure 245, lane 8.

10 GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 188 (lane 13; MW 92kDa). Purified GBS317dC-GST is shown in Figure 245, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1376

15 A DNA sequence (GBSx1461) was identified in *S.agalactiae* <SEQ ID 4213> which encodes the amino acid sequence <SEQ ID 4214>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -11.94    Transmembrane    132 - 148 ( 123 - 156)
    INTEGRAL    Likelihood = -11.09    Transmembrane    190 - 206 ( 183 - 209)
20    INTEGRAL    Likelihood = -4.94    Transmembrane    95 - 111 ( 94 - 115)

----- Final Results -----
          bacterial membrane --- Certainty=0.5776 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
25          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related sequence was also identified in GAS <SEQ ID 9133> which encodes the amino acid sequence <SEQ ID 9134>. Analysis of this protein sequence reveals the following:

```

30    Possible site: 22

>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -7.32    Transmembrane    126 - 142
    INTEGRAL    Likelihood = -6.90    Transmembrane    178 - 194
35

----- Final Results -----
          bacterial membrane --- Certainty=0.3930 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
40          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 94/204 (46%), Positives = 139/204 (68%)

Query: 5    LMKDKLLVVLTWIWIISLATLATIYIWLVIYPIRQFLKLEKVVYLKATIIYVFNKLM 64
          +M + + +W+H++LA L TIY WL YP+E+ LKLE+VV++ + I +N+N L+
45    Sbjct: 4    VVVENTELKSNVWVLATAILITITYSTWLYPFLVDHKLKQVVFMSKDAIHLNYNGLLN 63

Query: 65    YLTHPFISDLNMPSPFSSDGLKHFADVKYFLPTLAHGLFVLIITFPVLYFLRRGWKQKSI 124
          YLT+PF++ L +F SS DGLKHFADVK+LF L +F+ L +P + + K K +
50    Sbjct: 64    YLTNPFVITRLFPANFHSSADGLKHFADVKVLPHTLQVVLGLLIYPTKTFITQLKATKRFW 123

```

-1508-

Query: 125 LYEGFFKIAINLPTFIVVCARLLGFDQFFTLFHEVLFPQDSTWQENPLDTPVWILPETP 184  
 L + +A + P+ I + A +GF+ FFTLEH+VLF GDS+W F+PL D VINILPE F  
 Sbjct: 124 LLQKPLLAALFPIMIGLMASTIGFERHFFTLFHVQLFVGDSSMLFDLKDLSVINILPEVF 183

Query: 185 FLHCFIPLLIYETITILLIIGR 208  
 FLHCF+ F+++YE I L+ + R  
 Sbjct: 184 FLHCFLFRMIVYETIILWSLGLAR 207

- 10 SEQ ID 4214 (GBS167) was expressed in and purified from *E.coli*. The purified protein is shown in lanes 5 & 6 of Figure 223.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1377

- 15 A DNA sequence (GBSx1462) was identified in *S.agalactiae* <SEQ ID 4217> which encodes the amino acid sequence <SEQ ID 4218>. This protein is predicted to be p-nitrophenyl phosphatase (pho2). Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence

- 20 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3925 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15219 GB:Z99120 similar to N-acetyl-glucosamine catabolism  
 [Bacillus subtilis]  
 Identities = 121/249 (48%), Positives = 172/249 (68%)

- 30 Query: 3 YKGYLIDLGGTYKGSRI FAGERPIERLQEGGIPYMLVTNNITPTESVQEMLRGFNVE 62  
 YKGYLIDLGGT+Y G +I F+ L+++G+PY+ VTN++RTP+ V + L F++  
 Sbjct: 4 YKGYLIDLGGTYMGTEKIEEACEFVRTLKDRGVPLFVTVNRSRTPEQVADKLVSFDP 63

- 35 Query: 63 TPLETIYTAIMATVDYMNDRGKTAYVIGERGLKKAIDAGVEDITNPNPAYVVVGLDWN 122  
 E ++T +MAT ++ + + YVIGEBG+++AI + G +N +VVVG+D +  
 Sbjct: 64 ATECVQFTSMATAQHIAQQKIDASVVVIGESGIRQAIERNGLTFGRNADFVVVGLDWS 123

- 40 Query: 123 VTYDKLATATIAIQNGALFIGTNPDLNIPTERGLLPGAGSLNALLRAATRIKPFVFIGKPN 182  
 +TY+K A LAI+NGA FI TN D+ IPTERGLLPG GSL ++L +T ++PVFIGKP  
 Sbjct: 124 ITRYKFAVGCIAIRNGARFISTNGDIAIPTERGLLPGNGSLTSLVTVSTGVQVPFIGKPE 183

- Query: 183 AIIMNKALEILINIPRNQAVMVGVNYLTIDIMAGINNDITLLAVTGTFTVESVPDLPIQPS 242  
 +IIM +A+ +L ++ +MVGVNY TIDIMAGIN +DTLLV TG T E + D +P+  
 45 Sbjct: 184 SLIMEQMRVLGTVDSSETLMGVENYATIDIMAGINAGDTLLVHTGVTKREHMTDDMEKPT 243

Query: 243 YVLASLDWE 251  
 + + SL EW  
 Sbjct: 244 HAIDSLEWE 252

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4219> which encodes the amino acid sequence <SEQ ID 4220>. Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.53 Transmembrane 128 - 144 ( 128 - 144)

- 55 ----- Final Results -----  
                   bacterial membrane --- Certainty=0.1213 (Affirmative) < succ>

-1509-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP: CAB15219 GB: Z99120 similar to N-acetyl-glucosamine catabolism  
 [Bacillus subtilis]  
 Identities = 121/250 (48%), Positives = 166/250 (66%), Gaps = 1/250 (0%)

10 Query: 3 YKGYLIDLDGTYQGKRI PAGERFPIKRLQERGI PYLLVTNNITRTPEMVQSMLANQPHV 62  
 YKGYLIDLDGT+Y G +I F++ L++RG+PYL VTN++RTP+ V L + F +  
 Sbjct: 4 YKGYLIDLDGTYMGTEKIBACEFPVRLTGORGVPYLFVTNNSSRTFKQVADKLVS-FDI 62

Query: 63 ETSIETIYTATMATVDYNDMMRGKTAYVIGETGLKSAIAAAGYVEELENPAYVVVGLDS 122  
 + E ++T +MAT ++ + + YVIGE G++ AI G EN +VVVG+D  
 15 Sbjct: 63 PATEEQVFTTSMATAQHIAQKKDASVYVIGEGRIQALIEENGLTFGGNADFPVVVGIDR 122

Query: 123 QVTYEMLAIALTAIQGALFIGTNPDLNIPTERGLMPGAGALNALLEAATRVKPVFIGKP 182  
 +TYE A+ LAI+ GA FI TN D+ IPTERGL+PG G+L ++L +T V+PVFIGKP  
 20 Sbjct: 123 SITYEKFAVGCLAIRNGARFISINGDIAIPTERGLPGNGSLTSLVLTSTGVQPVFIGKP 182

Query: 183 NAIIMNKSLEVLGIQRSEAVMVGNLYTDIMAGI QNDIATILVTGTPEEVPFLFIQ 242  
 +IIM +++ VLG SE +MVGDNY TDIMAGI + T+LV TG T+ E + +P  
 Sbjct: 183 ESTIIMQAMRVLTGDVSETLMVGDNYATDIMAGINAGMDTLLVHTGVTKREHMTDMEKP 242

25 Query: 243 DRVLSLDEW 252  
 H + SL EW  
 Sbjct: 243 THAIDSLTEW 252

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 207/250 (82%), Positives = 227/250 (90%), Gaps = 1/250 (0%)

Query: 3 YKGYLIDLDGTYGKSRIPAGERFIERLQKGI PYMLVTNNITRTPEMVQSMLRG-FNV 61  
 YKGYLIDLDGTIY+GK+RIPAGERFI+RLQE+GIPY+LVNTNITRTE VO ML F+V  
 35 Sbjct: 3 YKGYLIDLDGTYQGKRI PAGERFPIKRLQERGI PYLLVTNNITRTPEMVQSMLANQPHV 62

Query: 62 ETPLETIYTATMATVDYNDMMRGKTAYVIGEGGLKKAIDAGYVEDTNPAYVVVGLDW 121  
 ET +ETIYTATMATVDYNDMMRGKTAYVIGE GLK AIA AGYVE+ +NPAYVVVGLD  
 Sbjct: 63 ETSIETIYTATMATVDYNDMMRGKTAYVIGETGLKSAIAAAGYVEELENPAYVVVGLDS 122

40 Query: 122 NVTYDKLATATLAIQNGALFIGTNPDLNIPTERGLPGAGSALNALLEAATRIKPVFIGKP 181  
 VTY+ LA ATLAIQ GALTFIGTNPDLNIPTERGL+PGAG+LALLEAAT+KPVFIGKP  
 Sbjct: 123 QVTYEMLAIALTAIQGALFIGTNPDLNIPTERGLMPGAGALNALLEAATRVKPVFIGKP 182

45 Query: 182 NAIIMNKALEILNIPRQAVMVGNLYTDIMAGINNDIITLLVTGTFTVEEVPDLFIQ 241  
 NAIIMN+LE+L I R++AVMVGNLYTDIMAGI NDI T+LVTTGFT EEVP LPIQ  
 Sbjct: 183 NAIIMNKSLEVLGIQRSEAVMVGNLYTDIMAGI QNDIATILVTGTPEEVPFLFIQ 242

Query: 242 SYVLASLDEW 251  
 +VL+SLDEW  
 50 Sbjct: 243 DRVLSLDEW 252

A similar DNA sequence was identified in *S. pyogenes* <SEQ ID 4215> which encodes amino acid sequence <SEQ ID 4216>. An alignment of the GAS and GBS sequences follows:

Identities = 94/204 (46%), Positives = 139/204 (68%)

55 Query: 4 VMVENIKLLCSMWILLALAILITTYSTWLNWYPLEVDHKLQGVVFMKDAIIMNYNGLIN 63  
 +M + ++ +W+N+++LA L TIY WL YP+E+ LKLE+VV++ + I +N+L+  
 Sbjct: 5 LMKDKLLVWITWIIWISLATIATYIWLNIYPIEQFLKELKVVLKASTIYFNENKMI 64

60 Query: 64 YLTNPFPVTRLEFANFESSADGLKHADVKMLFHLTQVVFGLLYLTPTKFTFQRLKTRFW 123  
 YLT+PF++ L +P SS DGLKHADVK+LF L +F+ L +P + + K K +  
 Sbjct: 65 YLTHPFTSIDLMPSPSSSEDGLKHADVKYLFTHLHGLFVILTFPVLYFLERGMQKKSIF 124

Query: 124 LLQKPLILAAALFPLMIGMASPIGFHFHTFLHQVLFVGDSSNLFDPLDKSDVIMILPEVF 183

-1510-

L + +A + P+ I + A +GF+ PFTLPH+VLF GDS+W P+PL D VINILPE F  
 Sbjct: 125 LYEGFFKIAIMLPIPIVVCAPILGFDQFFTLFHEVLFGDSTWQFNPLTDPVINILPETF 184

Query: 104 FLHCFLEFMIVVRIIIMSLVGLAR 207  
 FLHCF+ F+++YE I L+ + R  
 Sbjct: 185 FLHCFIPLIIVETITILLIIGR 208

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 10 Example 1378

A DNA sequence (GBSx1463) was identified in *S.agalactiae* <SEQ ID 4221> which encodes the amino acid sequence <SEQ ID 4222>. This protein is predicted to be oleoyl-acyl carrier protein thioesterase. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3332 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA02069 GB:BA026647 acyl carrier protein thioesterase  
 [Arabidopsis thaliana]  
 Identities = 59/248 (23%), Positives = 104/248 (41%), Gaps = 30/248 (12%)

Query: 2 GLLYRSTYEVPTFYESDTHMYMLPQLALALQISAKQSLKLGIDG-----LVFKRYGLV 56  
 GL Y+E + V YE +N + + L ++ + +G D ++ L+

Sbjct: 81 GLSYKSKFVRSYEVGSGNKATVETIANLLQEVGCHQAQSGVGFSTDFATTIMRKLHLI 140

Query: 57 WVVDYIIDIERLPKHAKIVILETEAKARNKLLCYRYFYIYGE-DGQKIITISSAFVLMD 115  
 WV I+I + P + + IET ++ ++ R + + G+ +S +V+M+

Sbjct: 141 NVIARMHIEIYKYPANGDVVEIETWCQSEGRIGTRRDWILKDSVTGEVIGRATSKVMMN 200

Query: 116 FKTRKIHPLVDDITSY-----QSQRHKVIRGPKYHPIGDSKVQYHVR 160  
 TR++ V DD+ Y ++ +KK+ PK + R

Sbjct: 201 QDTRRLQKVSODVRDEYLVFCQEPRLAFFENRNSLKKI---PKLEDPAQYSIMGLKPR 257

Query: 161 YFDLDMNGHVNSKYLEWYDVLDDPLSSHIPKKIDLKYKEIQYGTDIKSHMYQDGLV 220  
 DLDNM HVNN Y+ W+ + + + +H + I L Y +E Q + D L

Sbjct: 258 RADLDMNGHVNNVTYIGWVLESIPQETVDTHQLQVITLDYRRCQQDDVV-----DSLIT 311

Query: 221 TRHDIIGG 228  
 T IGG

Sbjct: 312 TTTSEIIGG 319

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4223> which encodes the amino acid sequence <SEQ ID 4224>. Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.88 Transmembrane 21 - 37 ( 21 - 38)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2550 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:



-1511-

>GP:AA571730 GB:U65643 acyl-ACP thioesterase [Myristica fragrans]  
 Identities = 41/128 (32%), Positives = 67/128 (52%), Gaps = 11/128 (8%)

Query: 33 PIFMIKRGGLVLDILAYFALLNPDTRKVTATIPEDLVAPFETDPVKKIHVR----PKMPL 87  
 F+ K G +L ++ +N TR+++ IPE++ E FV+ H V K+P  
 Sbjct: 147 FLRDKCTGKILTRATSVVWMMNKKTRLSKIPFVRVKEIFYPV--HGVLDSESKLKP 204

Query: 68 LEQS----IDEDYYRYFYDIDMNGHVNNSKYLDMNYDLGCEFLKTHQPLKMTLKCVKEV 143  
 L + I R + R+ D+D+N HVN+ KY+ W+ + + L+H+ MTL+Y KE  
 Sbjct: 205 LNDNTANYIRRGLAFRWSDLDVNQHVNNVKYIGWILESVPSLLESHLYGTMTEYRKEC 264

Query: 144 SPGGQILTS 151  
 G + S  
 Sbjct: 265 GKDGLLS 272

An alignment of the GAS and GBS proteins is shown below.

Identities = 62/144 (43%), Positives = 94/144 (65%)

Query: 101 GQKIITISSAFVIMDFKTRKIHFPVLDITSIYQSORIKKVIKRGPKYHPIGDSKVKQYHVR 160  
 G ++ I + F L++ TRK+ + +D+ + ++ +KK+ R PK + S + Y+VR  
 Sbjct: 40 GGLLVLDILAYFALLNPDTRKVTATIPEDLVAPFETDPVKKIHVRPKMPLQGSIDRYYVR 99

Query: 161 YFDLDMNGHVNNSKYLDMNYDLVLDLFLSSHPKKIDMKYKEIQYGTIDKSHVQDGLV 220  
 YFD+DMNGHVNNSKYL+WMYDVL +FL +H P K+ LKY+KE+ G I S ++ D L  
 Sbjct: 100 YFDIDMNGHVNNSKYLDMNYDLVGLCEFLKTHQPLMTLKVKVRESFGQITSSYHLDQLT 159

Query: 221 TRHDIGGDAIHQAARIEWQEKKE 244  
 + H I ++AQA IEW+ K+  
 Sbjct: 160 SYHQITSDGQLNAQAMIEWRAIKQ 183

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1379

A DNA sequence (GBSx1464) was identified in *S. agalactiae* <SEQ ID 4225> which encodes the amino acid sequence <SEQ ID 4226>. This protein is predicted to be coproporphyrinogen III oxidase. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1484 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05062 GB:AP001511 coproporphyrinogen III oxidase [Bacillus halodurans]  
 Identities = 173/375 (46%), Positives = 248/375 (66%), Gaps = 5/375 (1%)

Query: 5 PTAAXVHIIPFCTQICYCDFSKVPIKQPVDAYLQALIREER----SYDITSLRTLYIGG 60  
 P +AY+HIIPFC ICYCDF+K ++KNQF+ YLQAL E L+TL+AG  
 Sbjct: 2 PKAAYIHIPFCEHCYCDFPKFKLNQPVNEYLQALTEEMAMVVVARQPKSLQITLYVGG 61

Query: 61 GTPTSIASVOLDYLLTSLSDRLNLTLEETPIRANPGDITVDKIEVLQKSANRVSLGVQ 120  
 GTPT+AA QL LL + R L L+ LEEPT E NP + +K++VL+ V+R+S+GVQ  
 Sbjct: 62 GTPTALTADQALQALLSKRTLPLSLDEETPEVNPDSIDEEKLDVLRSGVDRLSIGVQ 121

Query: 121 TFNDKHLKKIGRSHNEAQIYSTIDALCTAGFQNISLDLIYALPGQITMDVRSNVAKLSL 180  
 F LK IGR+H++ + ++ +AGF N+S+DL+ LP QT + +A +L  
 Sbjct: 122 AFQPLLKKEIGRTHDQKEVQAVEKSRQAGFANLSLDLMLGLPKQTFEMPAETLKRAFAL 181

-1512-

Query: 181 NIPHLISLYSILEHHTVFMNMNRGKHLPTDLEAEFYEIISMERNGFHEVEISNPT 240  
 + HLS YSL +E TVF N+ R+G+L LP ED E +N+ + E E+GF+ YEISNF  
 Sbjct: 182 EVELHSCYSLKVEAKTVFNMNRGKGLPTDLEAEFYEIISMERNGFHEVEISNPT 241

5 Query: 241 KPGFESRHNLMYWNVEYVYGVGAGASYLDGIRYRNNGPIQHVLKGVSGNARLSR-EVL 299  
 K G+BSRHNLMYWNVEYVYGVGAGASYLDGIRYRNNGPIQHVLKGVSGNARLSR-EVL  
 Sbjct: 242 KPGYESRHNLMYWNVEYVYGVGAGASYLDGIRYRNNGPIQHVLKGVSGNARLSR-EVL 301

10 Query: 300 SKNEMMBEELFLGLARKKBSVIGKFEQKPGTSPEKRYGQIVQSLQSDGLLKNNQFICMT 359  
 S+ E MEE++FLGLARK+ GV F ++FG S Y + + +L+ LL+ + +++T  
 Sbjct: 302 SRVCEMBQMPFLGLARKKBSVIGKFEQKPGTSPEKRYGQIVQSLQSDGLLKNNQFICMT 361

Query: 360 KGLFLGDTVAEKFI 374  
 +GL LG+ V E+F+  
 15 Sbjct: 362 DGLLLGNEVFEQL 376

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4227> which encodes the amino acid sequence <SEQ ID 4228>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3202 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 304/376 (80%), Positives = 343/376 (90%)

30 Query: 1 MLKKPTSAIVHIPFCTQICYICDFSKVFIKQPVDAVLQALIREPSYDITELTYLIGG 60  
 M KKPTSAIVHIPFCTQICYICDFSKVFIKQPVDAVLQALIREPSYDITELTYLIGG  
 Sbjct: 33 MSKKPTSAIVHIPFCTQICYICDFSKVFIKQPVDAVLQALIREPSYDITELTYLIGG 92

35 Query: 61 GTPTISAVQLDYLLTELSRLNLTLESFTIEANPGDLTVDAKIEVLQKSAVNRSLGVQ 120  
 GTPTI+I+ QL+YLL L R+TNI+ LEEPTIEANPGDLT +KI VLQ+SAVNR+SLGVQ  
 Sbjct: 93 GTPTAITAQLEYLLNHLERNLMDLESFTIEANPGDLTPEKIAVLQKSAVNRSLGVQ 152

Query: 121 TFDNKHLEKRGSHNEAQIYSTIDALKTAGFQINISIDLIALPGQTMDDVRNVAKALSL 180  
 TFN+K LK+IGSHNE QIYSTI LKTAGF NISIDLIALPGQTMDDVRNVAKALSL  
 40 Sbjct: 153 TFDNKHLEKRGSHNEAQIYSTIDALKTAGFQINISIDLIALPGQTMDDVRNVAKALSL 212

Query: 181 NIPHLISLYSILEHHTVFMNMNRGKHLPTDLEAEFYEIISMERNGFHEVEISNPT 240  
 +IHLISLYSILEHHTVFMNMNRGKHLPTDLEAEFYEIISMERNGFHEVEISNPT  
 Sbjct: 213 DIPHLISLYSILEHHTVFMNMNRGKHLPTDLEAEFYEIISMERNGFHEVEISNPT 272

45 Query: 241 KPGFESRHNLMYWNVEYVYGVGAGASYLDGIRYRNNGPIQHVLKGVSGNARLSR-EVL 300  
 KPGFESRHNLMYWNVEYVYGVGAGASYLDGIRYRNNGPIQHVLKGVSGNARLSR-EVL  
 Sbjct: 273 KPGFESRHNLMYWNVEYVYGVGAGASYLDGIRYRNNGPIQHVLKGVSGNARLSR-EVL 332

50 Query: 301 KNEEMMBEELFLGLARKKBSVIGKFEQKPGTSPEKRYGQIVQSLQSDGLLKNNQFICMT 360  
 K EMBEEMMBEELFLGLARKKBSVIGKFEQKPGTSPEKRYGQIVQSLQSDGLLKNNQFICMT  
 Sbjct: 333 KNEEMMBEELFLGLARKKBSVIGKFEQKPGTSPEKRYGQIVQSLQSDGLLKNNQFICMT 392

Query: 361 KGLFLGDTVAEKFI 376  
 KGLFLGD+VAE+FI+  
 55 Sbjct: 393 KGLFLGDVAERFILD 408

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1513-

**Example 1380**

A DNA sequence (GBSx1465) was identified in *S.agalactiae* <SEQ ID 4229> which encodes the amino acid sequence <SEQ ID 4230>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3729 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1381**

A DNA sequence (GBSx1466) was identified in *S.agalactiae* <SEQ ID 4231> which encodes the amino acid sequence <SEQ ID 4232>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2989 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4233> which encodes the amino acid sequence <SEQ ID 4234>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2993 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 36/109 (33%), Positives = 58/109 (53%), Gaps = 6/109 (5%)

Query: 9  WAGHKYLVLSFKSQKIYLDIRQ/ILKSPNCT---VLEVQSLIDQAVLLESPSQVFNAYMHI 65
      WA  XY  V++  SQ+  Y  +R+  K  +  VL  LI++A  +  +  +  AY  H+
Sbjct: 13  WATQKYVWMAFSQQHYNALRELFLKQNWSESKVLTFHCLTSEAQAIPPTVKSLEPTAYQHV 72

Query: 66  WGYFKNKAERQEKSEFLTLLEKYKTYGQRKLLAFLKQLLAKYPSNLY 114
      WGYFK  A  ++EK+  F  L  +  +  ++L  FL+++  A  Y  SYL
Sbjct: 73  WGYFKVKSQSEKDHFKDLDAQLLET---KSEEMLCFLQEWTAHYQPSTYL 118

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1382

A DNA sequence (GBSx1467) was identified in *S.agalactiae* <SEQ ID 4235> which encodes the amino acid sequence <SEQ ID 4236>. This protein is predicted to be mrsA (mrsA). Analysis of this protein sequence reveals the following:

```

5      Possible site: 35
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.96    Transmembrane    56 - 72 ( 56 - 72)

10      ----- Final Results -----
      bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:CAB11970 GB:Z99105 similar to phosphoglucosylase (glycolysis)
      [Bacillus subtilis]
      Identities = 284/451 (62%), Positives = 353/451 (77%), Gaps = 4/451 (0%)

20      Query: 1   MGKYPGTGVRGEANVELTPELAFKLRFGGYVLSQHETDRPRVVFVARDTRISGEMIESA 60
      MGKYPGTGVRG AN ELTPELAFKLRFGGYVLSQHETDRPRVVFVARDTRISGEMIESA 60
      Sbjct: 1   MGKYPGTGVRGVAENSELTPELAFKVRFGGYVLTG-DKQRPKVLIGRDRTRISGHELEA 59

      Query: 61   LIAGLLSGVIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFPGSDGFKL 120
      L+AGLLSG+EV +LGV++TPGVSYL + A AGVMISASHNP DNGIKFPG DGFKL
      Sbjct: 60   LVAGLLSGVIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFPGSDGFKL 119

      Query: 121  DDRELEIEAILDAKEDTLPRPSAQGLTLDVDPBGLRKYKPMESTGI-DLBMKVALD 179
      D++E EIE L+D ED LPRP LG + DY BG +KY +F++ T D G+ VALD
      Sbjct: 120  SDSQFARIETRMDEPRDKLPRVCGADLGVNDYVFGQKYLQLFKTADEFTGTHVALD 179

30      Query: 180  TANGAATASARNIFLDIADISVIGDQPDGLNINDGVGSTRPEQLQSLVRENGSDIGLAF 239
      ANGA ++ A ++F DL+AD+S +G P+GLNINDGVGSTRPE L + V+E +D+GLAF
      Sbjct: 180  CANGATSSLAHLFADLDAOVSTWGTSPNGLNINDGVGSTRPEALSAFVKKKNADIGLAF 239

35      Query: 240  DGDSDRLIAVDENGETVDGDKIMFIIGKYLSDKGQLAQNTIVTTVMNSLGFHKAIDREGI 299
      DGD DRLIAVDG G IVDGD+IM+I K+L +G+L +T+V+TVMSNLGF+KAL++EGI
      Sbjct: 240  DGDGDRLIAVDENGINVDGQIMYICSKHLKSEGRKDDTVVSTVMNSLGFYKALESGEI 299

40      Query: 300  HKAITAVGDRYVVEEMRKSGYNLGGSGHVIIMDYNITGQGQLTAIQLTKVMKSTGKGL 359
      TAVGDRYVVE N+K GYN+GSGSGH+I +DYNITGQG L+AI L +G TGR L
      Sbjct: 300  KSVQTAVGDRYVVEAMKDGYNVGGSGHGLI FLDYNITGQGLLEAIDMLNTLQATGKFL 359

      Query: 360  SELASEVTIYPQKLNVIRVENNMKDKAMEVPAIARIABKMEEMDNGRIIVRPSGTPEFL 419
      SELA+E+ +PQ LVN+RV + K K E + +I+++E+EM+G+GRILVRPSGTPEFL
45      Sbjct: 360  SELAAEMQKFPQLLVNVRVTD--KYKVEENKVKAVISEVEKEMNGDRIIVRPSGTPEFL 417

      Query: 420  LRVMAEAPTEAVDYVDTIADVVRTEIGLD 450
      +RVMAEA T E D YV+ I +VVR+E+GL+
      Sbjct: 418  VRVMAEAKTKELDSTYVNRIVVEVVRSENGLE 448

50

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4237> which encodes the amino acid sequence <SEQ ID 4238>. Analysis of this protein sequence reveals the following:

```

55      Possible site: 35
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.96    Transmembrane    56 - 72 ( 56 - 72)

60      ----- Final Results -----
      bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1515-

The protein has homology with the following sequences in the databases:

>GP:CB11970 GB:Z99105 similar to phosphoglucomutase (glycolysis)  
[Bacillus subtilis]

Identities = 287/451 (63%), Positives = 346/451 (76%), Gaps = 4/451 (0%)

5 Query: 1 MGKYPGTDGVRGEANVELTPELAFKLRPGGYVLSQHETERPKVFPVARDTRISGEMLESA 60  
Sbjct: 1 MGKYPGTDGVRGEANVELTPELAFKWRPGGYVLTG-KDQRPKVLIGRDTRISGEMLEGA 59

10 Query: 61 LIAGLLSGVIEYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGSDGFKL 120  
Sbjct: 60 LVAGLLSIGAEVRLGVISTPGVSYLTAMADAGVMISASHNPVQDNGIKFFGSDGFKL 119

15 Query: 121 ADDQRELEIKALLDAPEDTLPRPSAAGLTGLVDYFPGRLRYEKFVLTGTG-DLSKMTVALD 179  
Sbjct: 120 SDQREAEIERLADPEFDKLPFRFGADLGVDYFPGGQKYLQFLKQTADEDTGIRVALD 179

20 Query: 180 TANGAASVSARDVFLDLNABIAVIGEKPKWLNINDGVGSTRPEQLQELVKETGADLGLAF 239  
Sbjct: 180 CANGATSSSLATHLFDLADVSTMTGSPNGLNINDGVGSTRPEALSFAVKEKNADLGLAF 239

25 Query: 240 DGDSDRLIAVDDEGRIVDGRIMFLIGKYLSEKLLAHNTVITVTMSNLGPHKALDKQGI 299  
Sbjct: 240 DGDSDRLIAVDKSGNIVDGGIMYICSKHLSEKRLADTVTSTVMSNLGPHKALDKQGI 299

30 Query: 300 NKAITAVGDRYVVEBMRSGSYNLGSGSGHVIIMDYNTTGDQLTAIQLAKVKETGKSL 359  
Sbjct: 300 KSVQTAVGDRYVVEAMKDGYNVSGSGSHLIFLDYNTTGDQLTAIQLAKVKETGKSL 359

35 Query: 360 SELAAEVITYPQKLVNIRVENMICEKRAHEVPAIANIADMBEDMAQKRLIVRPSGTEPL 419  
Sbjct: 360 SELAAEMQKFPQLLVNRVVD--KYKVEENKVKAVISEVEKMGDGRILVRPSGTEPL 417

40 Query: 420 LRVMAEAPTDAEVDYVYVDITADVVRTEIGCD 450  
Sbjct: 418 VRVMAEAKTEKLCDEYVNRIVEVVRSEMGLE 448

An alignment of the GAS and GBS proteins is shown below.

Identities = 400/450 (88%), Positives = 429/450 (94%)

40 Query: 1 MGKYPGTDGVRGEANVELTPELAFKLRPGGYVLSQHETDEPRVFPVARDTRISGEMLESA 60  
Sbjct: 1 MGKYPGTDGVRGEANVELTPELAFKLRPGGYVLSQHET-EP-VFPVARDTRISGEMLESA 60

45 Query: 61 LIAGLLSGVIEYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGSDGFKL 120  
Sbjct: 61 LIAGLLSGVIEYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGSDGFKL 120

50 Query: 121 DDRELEIEALLDAKEDTLPRPSAAGLTGLVDYFPGRLRYEKFMTSTGIDLEKMKVALDT 180  
Sbjct: 121 ADDQRELEIEALLDAPEDTLPRPSAAGLTGLVDYFPGRLRYEKFVLTGTG-DLSKMTVALDT 180

55 Query: 181 ANGAATASARNIFLDLNDASIVIGDQDGLNINDGVGSTRPEQLQSLVRENGSDIGLAFD 240  
Sbjct: 181 ANGAASVARDVFLDLNABIAVIGEKPKWLNINDGVGSTRPEQLQSLVRENGSDIGLAFD 240

60 Query: 241 GDSDRLIAVDENGRIVDGRIMFLIGKYLSDKQQLAQTIVITVTMSNLGPHKALDRGSIH 300  
Sbjct: 241 GDSDRLIAVDGTEGIVDGRIMFLIGKYLSEKLLAHNTVITVTMSNLGPHKALDRGSIH 300

65 Query: 301 KAITAVGDRYVVEBMRSGSYNLGSGSGHVIIMDYNTTGDQLTAIQLAKVKETGKSL 360  
Sbjct: 301 KAITAVGDRYVVEBMRSGSYNLGSGSGHVIIMDYNTTGDQLTAIQLAKVKETGKSL 360

Query: 361 ELASEVITYPQKLVNIRVENMICEKRAHEVPAIANIADMBEDMAQKRLIVRPSGTEPL 420  
Sbjct: 361 ELASEVITYPQKLVNIRVENMICEKRAHEVPAIANIADMBEDMAQKRLIVRPSGTEPL 420

-1516-

Sbjct: 361 ELAAEVTIYPQKLVNIRVENSMSKERAMEVPALANIYAKMEDEMAGNGRILVVRPSGTEPLI 420

Query: 421 RVMAEAPTNEAVDYVDTIADVVRTIGLD 450

RVMAEAPT+ VDYVDTIADVVRTIG D

Sbjct: 421 RVMAEAPTDAEVDYVDTIADVVRTIGCD 450

SEQ ID 4236 (GBS402) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 5; MW 78kDa).

GBS402-GST was purified as shown in Figure 218, lane 3-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1383

A DNA sequence (GBSx1468) was identified in *S. agalactiae* <SEQ ID 4239> which encodes the amino acid sequence <SEQ ID 4240>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11969 GB:Z99105 ybbr [Bacillus subtilis]

Identities = 90/324 (27%), Positives = 167/324 (50%), Gaps = 18/324 (5%)

Query: 1 MKKFFTNKFLGVSLFLAILLFLTATATSMHQNSKIAG-----ASEYTHLTLDVPI 55

M KF N++ + +++L A+L++ A + N K G S T TLTD+P+

Sbjct: 1 MDKFLNIRWAKVLIALLFALLLYV---AVNSNQAPTKKPGSEFFPTSTDEATLTDIPV 57

Query: 56 DIKYSDDYFISGYSGADVMS-SVNRKLDSEINEDTRKFKVADLTNRKPTHKVPL 114

YD ++Y ++G +Y + S + VK + T+ F++ AD+ ++K GTHK L

Sbjct: 58 KAYIDDERVVTGVPQTWVVTIKGSTSAVKARQ---TKNFIYADMRLKGTGTHKVEL 113

Query: 115 KVNLPSSGVNATVSPPTTITVMKKKKTKPEPV-YGVNDKQIKGAVNDMSVDVSKVKV 173

K N+ G+ +++P+ TPT+ ++ TK PPV + N ++K GY+ + V V++

Sbjct: 114 KARNVSDGLTISINPSVTTVTIQRITKSPFPVEVETNKSMMKKGSPEQPIVSPKVNQI 173

Query: 174 TSDESIIDRIDHVAANI PDKVLDODFNKIVTLCVATADGTIVASIIHPSKATLSVKVKV 233

T +++ID I A++ + D+ K + DG L + PS ++V V

Sbjct: 174 TGSKNVIDNISLHKASVNLNA-DETIEKAKVTVYDKGNALFVDVEPSVIKITVPVTS 232

Query: 234 LKTVPINILPVGGSDSISKINYKLSQKAVISGTSKEALEAISVIN-AEVDISDVTKPT 292

+K VP + G D + S N + S + + G++ L++ I+ +D+S + K+

Sbjct: 233 PSKKVPFKIERTGSLFDGVSIDNIESPSEVTVYGSQDLVLEFDIGVSLDKINKDS 292

Query: 293 --EKKINLSANNVSVDPQVTVQL 314

E I L + P++VT+ +

Sbjct: 293 DIEDADIPLKGVKKISPSKVTPLHI 316

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4241> which encodes the amino acid sequence <SEQ ID 4242>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

-1517-

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP: CAB11969 GB:Z99105 yobR [Bacillus subtilis]  
 Identities = 81/322 (25%), Positives = 154/322 (47%), Gaps = 15/322 (4%)

10 Query: 1 MKRFLNRPWLGWVVFPAILLFLTAASNNH---NNSSQIYSPIETHTSLKDVPIDM 56  
 M +FLN+R +++++ FA+LL++ A +SN + T +L D++P  
 Sbjct: 1 MDKFLNRPWLVKIALLLFALLLV-ANVNNQAPTPKKPGBSFFFTSTTDEATLTDIPVKA 59

Query: 57 KYDSKYPISGYSYGAEVYLT-STNRKLDSEVNNDRNFKIVADLTHSHPGTVSNLRV 115  
 YD + Y +G V + ST+ +K + T+NP+I AD+ H GT V L+  
 15 Sbjct: 60 YYDDENYVVTGVQPTVNVITKGSTSAVKKAQKQ---TKNFEIYADMEHLKTGTIRKVLKA 115

Query: 116 ENLPSGVYATVSPDKISVTIGKKESKVFVRGS-VDAKQIANGYEISKIETGVNKRVEIS 174  
 +N+ G+T +++P +VTI ++ +K PFV + ++ GY + V++T  
 20 Sbjct: 116 KNVSDGLTISINPSTVTTITQERTTKSFVEVEYNNKMKKGYSPPQPIVSPKNVOITG 175

Query: 175 DESTIALIDHVAKLDDQVLDNRNYSRVTLQAVSADOTILASADIPAKTNLSVAKKIT 234  
 ++ I I A + + D + DG L ++P+ ++V V +  
 Sbjct: 176 SKQVINISILHKASVNLNA-DETIEKAKVTYVDKGNALPVDVVEFSVKITITVPTVPS 234

25 Query: 235 KSVPIRVEAVGMMDDSLSDIQYKLSKQTAVISGSRVLEDIEDII-AEVNISDVTKT-- 291  
 K VP ++E G + D +S + S + GS++VI+ ++ I +++S + K++  
 Sbjct: 235 KKVPFKIERTGSLPDGVSIAINTSSPSEVTYVGSQVLDLSLEFIDGVSLDLSKINKSDSI 294

Query: 292 SKTVSLSSSQSIEPSSVTVQL 313  
 + L I PS VT+ +  
 30 Sbjct: 295 EADIPLDGVKIKSPSKVTLHI 316

An alignment of the GAS and GBS proteins is shown below.

Identities = 198/319 (62%), Positives = 251/319 (78%), Gaps = 1/319 (0%)

35 Query: 1 MKKFTTNKTLGVSFLAILLFLTATATSMNHQDNSKIAGSETYTHLTDPVIDIKYD 60  
 MK+F ++ WLG+VS+F AILLFLTA A+S ++ +S+I ETTYH+L DVID+KYD  
 Sbjct: 1 MKRFLNRPWLGWVVFPAILLFLTA-ASSNNHNSSQIYSPIETHTSLKDVPIDMKYD 59

40 Query: 61 SDOYFISGYSYGVYSSVMNKLDSSEINEDTRKFKVADLTNNKFTGTHKVELKVNLP 120  
 SD YFISGYSYGA+VY++S NR+KLDSE+D NTR FK+VADLT+ PGT V L+V NLP  
 Sbjct: 60 SDKYFISGYSYGAEVYLTSTNRKLDSEVNNDRNFKIVADLTHSHPGTVSNLRVENVLP 119

45 Query: 121 SGVNATVSPFTTITVTMCKKTKKFPVYGHVNDKQIKAGYAVDMKSDVSKVKVTSDESII 180  
 SGV ATVP I+VT+GKK++K PFV G V+ KQI GY + K+ V+KV+VTSDES I  
 Sbjct: 120 SGVTATVSPDKISVTIGKIGSKVFPVRGSVDAKQIANGYEISKIETGVNKRVEVTSDESI 179

Query: 181 DRIDHVAANIIPDKVLDLDDPKNTVITLQAVTADGTVLASIIHPSKATLSVKVKLTKTVPI 240  
 IDHV A+PDD+VLD +++ VTIQAV+ADGT+LAS I P+K LSV VKK+TK+VPI  
 50 Sbjct: 180 ALIDHVAKLDDQVLDNRNYSRVITLQAVSADGTILASADIPAKTNLSVAVKKITKSVPI 239

Query: 241 NLIPVQGFSDSEIKNYKLSQEKAVISGTEKALEAISVINAEVDSIDVTQTEKKIKLSA 300  
 + VG DS+S I YKLS++ AVISG++E LE I I REV+ISDVTRNT K ++LS+  
 55 Sbjct: 240 RVRAVGMDDSLSDIQYKLSKQTAVISGSRVLEDIDRIARVNISDVTKNTSKTVLS 299

Query: 301 NNVSVDPAQVTVQLITTTKK 319  
 + VS++P+ VTQVLTITTK  
 Sbjct: 300 SQVSIKPSVTVQLITTTKK 318

60 SEQ ID 4240 (GBS99) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 6; MW 35.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 9; MW 60.7kDa).

-1518-

The GBS99-GST fusion product was purified (Figure 197, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 293), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1384

A DNA sequence (GBSx1469) was identified in *S. agalactiae* <SEQ ID 4243> which encodes the amino acid sequence <SEQ ID 4244>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0503 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1385

A DNA sequence (GBSx1470) was identified in *S. agalactiae* <SEQ ID 4245> which encodes the amino acid sequence <SEQ ID 4246>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.50 Transmembrane 20 - 36 ( 18 - 46)
INTEGRAL Likelihood = -7.64 Transmembrane 48 - 64 ( 42 - 68)
INTEGRAL Likelihood = -3.40 Transmembrane 80 - 96 ( 80 - 96)

----- Final Results -----
bacterial membrane --- Certainty=0.4800 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11968 GB:Z99105 alternate gene name: ybbQ-similar to
hypothetical proteins [Bacillus subtilis]
Identities = 125/253 (49%), Positives = 186/253 (73%), Gaps = 5/253 (1%)

Query: 27 MDIIIVAVILYKFIKALAGTKIMSLIQVILFIIIRPVSEWIGL/TTTTFLMQIVTYGVIL 86
+DI+V +YK I + GTK + L+G++ + +R S++GL+T+ +L+Q +T+G +
Sbjct: 16 VDILLWVYIYKILIMVIRGTRKAVQLKIGIVIVIVIRNASQVILGLSLQWLMQAITWCF 75

Query: 87 AGVILFAPEIRTKLEKFGRTPOLFTQRSQSSDE---KLVDALVKAVAYMSPRKICALIS 143
A ++IF PE+R LE+ GR F RS +E K ++A KA+ YM+ R+IGAL++
Sbjct: 76 AIIIIIPQFLRRALSQLGRG--FFSRGGTPVSEAGQKTRAITKAINYMAKRRIIGALIT 133

Query: 144 IERTTILQRYIATGPIADADISSELLINIFIPNTPLHDGAVIVKDKIATACSYLPLSES 203
IER + +YI TGIPL+A +SSELLINIFIPNTPLHDGAVI+K+ +IA A YLPLSES
Sbjct: 134 IERTTGMDYIETGPIAAKVSSELLINIFIPNTPLHDGAVIMKNNIAAACYLPLSES 193

Query: 204 SSISKEFGTRHRAAIGLSNSDALTVIVSEETGSGISVA/KGRFLHDLKDSFEAIRLQ 263
```



-1519-

ISKE GTRHRAA+G+SE +D+LT+IVSEETGG+SWA G+ +L++++ +L +  
 Sbjct: 194 PFISKELGTRHRAAVGISEVTDLSLTIVSEETGGVSVAKNGDLHRELTEALKEMLEARF 253

Query: 264 IQNQRENSKLAWY 276

+N + S WY

Sbjct: 254 KGNTRDTSSNRWY 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4247> which encodes the amino acid sequence <SEQ ID 4248>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -6.64 Transmembrane 20 - 36 ( 19 - 40)  
 INTEGRAL Likelihood = -6.21 Transmembrane 48 - 64 ( 47 - 68)  
 INTEGRAL Likelihood = -2.07 Transmembrane 76 - 92 ( 76 - 92)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3654(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA003984 GB:AP001507 unknown conserved protein [Bacillus halodurans]  
 Identities = 117/255 (45%), Positives = 178/255 (68%), Gaps = 6/255 (2%)

Query: 19 PWL-LAVHLLDILVAYLYRPFKALGTGKIMSLVQGVIFPLVLRFAIEWGIFTTITILM 77

PWL +LDIL+V Y+Y+ I + GT+ + L++G+ L++ I+ + T+ ++

Sbjct: 8 PWNLYLTQILDLVVTYVYIKAMIIIRGTAVQLKLGITVILVYALISIFPLRLTGLNIV 67

Query: 78 NQVITYGVIAGVVITPEIRAGLEKPGRSTQVFLQKYVSSSAL- --VDALIKSVAYMS 134

NQ IYIG++A ++IF PE+R LE+ GR F + + E + +DA++K+ YMS

Sbjct: 68 NQAITYGLLAVIIIPQELRRALQLGRGR--FFASRTANEETIMKKTDAIVKASTYMS 125

Query: 135 PRKIGALIAIEQTQTLQBYIATGIPINADISSQLLINIFINPITPLMDGAVIGVQNKIVAA 194

R+IGALI++E+ + +Y+ TGIP+NA+++S+LLIN FIPITPLMDGAVI+ +I+AA

Sbjct: 126 KRRIGALISMERETGMDYVETGIPMAQNLTSLLINIFINPITPLMDGAVIINQITILAA 185

Query: 195 CAYLPLSESKAISKEPGTRHRAAIGLSENSDALTIIVSEETGAISVTRRQFLDLSTDE 254

YLPSE+ ISKE GTRHRAA+G+SE +D LTI+VSEETG IS+T+ G+ DL ++

Sbjct: 186 ACYLPLESNPFISKELGTRHRAALGVSEVTDCLTIIVSEETGHISLTINSELHRLDLEQ 245

Query: 255 FETVLRITYLMSNSNV 269

+L L+S + +

Sbjct: 246 LRSLEAELISEARN 260

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/283 (71%), Positives = 239/283 (84%), Gaps = 2/283 (0%)

Query: 1 MDIPSADSKFVASIMNPWILIHMDIIVAVLIVYKFKALGTGKIMSLQGVILFII 60

M+ S+ID KP S+ +PW++ +HL+DI+IVA LLY+FKAL GTKIMSL+QGVILFII 60

Sbjct: 1 MNLSLSDIKFLSLFADEFWLLAVHLLDILVAYLYRPFKALGTGKIMSLVQGVIFPLV 60

Query: 61 IRPVSSEWGLTITITIFMNVQIVTYGVIAQGVVFAPEIRITGLEKPGRTQLPTQRSQSSDE 120

+RF++EWIG TTT+LWQ+ +TYGVIAQGVVIF PEIR GLEKPGR+ Q+F Q+ +SS+

Sbjct: 61 LRFIAEWIGFTITITILMNVQIVTYGVIAQGVVITPEIRAGLEKPGRSTQVFLQKYVSSS 120

Query: 121 KLVDALVKAVAYMSPRKIGALISERTQTLQBYIATGIPLDADISSELLINIFINPITPLH 180

LVDAI+K+VAYM PRKIGALI+IR+TQTLQBYIATGIFL+ADISS+LLINIFINPITPLH

Sbjct: 121 ALVDALIKSVAYMSPRKIGALIAIEQTQTLQBYIATGIPINADISSQLLINIFINPITPLH 180

Query: 181 DGAVIVDKKIIATACSYLPLESSSISKEPGTRHRAAIGLSENSDALTIIVSEETGGISV 240

DGAVIV KI AC+YLPSE+ +ISKEPGTRHRAAIGLSENSDALTIIVSEETG ISV

Sbjct: 181 DGAVIVQNKIVACAYLPLESKAISKEPGTRHRAAIGLSENSDALTIIVSEETGAISV 240

Query: 241 ALKGEFLHLDLSDKSFALITQLIQNQRENSKLAWYNQLLRK 283

-1520-

EG+FLHDL S D FE +LRT L+ N N L WY ++L K  
 Sbjct: 241 TRGQFLHDLSTDEFETVLRITYLMSN--SNVTLFWYKILGGK 281

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1386

A DNA sequence (GBSx1471) was identified in *S. agalactiae* <SEQ ID 4249> which encodes the amino acid sequence <SEQ ID 4250>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.60 Transmembrane 33 - 49 ( 33 - 49)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1387

A DNA sequence (GBSx1472) was identified in *S. agalactiae* <SEQ ID 4251> which encodes the amino acid sequence <SEQ ID 4252>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1001(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9781> which encodes amino acid sequence <SEQ ID 9782> was also identified.

- The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC84012 GB:AF080002 UDP-N-acetylmuramyl tripeptide synthetase  
 MurC [Helicobacillus mobilis]  
 Identities = 143/442 (32%), Positives = 229/442 (51%), Gaps = 17/442 (3%)

Query: 12 GKSAYHLLSKWGRGST-YPGSLALKFDKIDLTIAKDYE--IVVVTGNGKTLTALTVG 68  
 GK+A +L + G G T +PG + + IL +A+ + +VVVTGNGKT T+ +  
 Sbjct: 2 GKTAIWLNRFGHGTSFPGGIGRRVAPQILTALARQLRGAMVVTGNGKTTTSKMLAA 61  
 Query: 69 ILKEAPGQVVTPNSGANMIGIVSTFLTAKSKSGC--KCIAVLRIDEASLPRTQVIKPS 126  
 I++++ + N +GN++ GI + F+ + + +E+DER++P++ ++P  
 Sbjct: 62 IVEKSSLLTTHNRAGANLVGGITTFADISATIGKSTSDIGILVDEBATIPQLVREVQPK 121  
 Query: 127 LFFVTNIFRDNDRYGEIYTTYQMILGGAANAP-QATILANGDSPLFNS--KSVTNPVQS 183  
 V TN FRDQ+DR+GR+ T ++ + P Q+ + N D PL S K V +  
 Sbjct: 122 GVVVTNFRDQLDRFGELDKTSLVIGALRLLPVQSIAVLNADDELVASLGKDFPQVLY 181  
 Query: 184 YGFNTDKHSPRLAHYNTGILCPKQAILTYRLNTYANLGDYTCNCDFERNPLDYALTR 243  
 +G + + R + E C C LTY + IG Y C +C FEP +T

-1521-

Sbjct: 182 FGIDRSYGAREMLQSAETFRCLRLOSHPLTYDMFFQQLGHYKCSHOGFERPEPKIKVIG 241

Query: 244 LTHLNTSSGPFVIDGQ----QYNINVGGLYNIYNALAAVVAEYFQVPSQIKDGFDKSR 299  
 + S F ++ Q ++ G YNINYNALAA++ A + I+ G R

5 Sbjct: 242 IQLKGGEGSAFTVETPRGTWQLSLSTPGFYNIYNALAAATASAIRLDPEKARAGLQGYR 301

Query: 300 AVFGRQETFTTIGN-KKCTLVLKKNFVGASQALDNKIAFYPFSLSVLLNANYADGIDTSW 358  
 PGR E + + ++ L LKNP G + + + P L V++N N ADG D SW

10 Sbjct: 302 TNFGMERILEDDGRAPLALIKMPTGCDENVIRTLVQNGPKRLVLIINDNADGRDLSW 361

Query: 359 IWDANFETI--LYMNIPEIFAGGVKHSIARLRVTVGYDEKRIK-QADKLQDITMIEQQ 415  
 +WDA+FE++ + + +F G+R ++A RL TG + I+ +A+ I + +E

15 Sbjct: 362 LWDADFSLELPVPEGRSVFTSGLRGSDMAIRLNTYTPASSIRYRANVESAIRSALEMT 421

Query: 416 ET-EHAYILATYTTAMLEFREIL 436  
 E E YIL TYTA+LE + L

15 Sbjct: 422 EPGETLYILPFTTALLESKAAL 443

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4253> which encodes the amino acid  
 20 sequence <SEQ ID 4254>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----  
 25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 343/446 (76%), Positives = 393/446 (87%)

Query: 1 MKINTALGVAAGKSAHYLLSKMGRSGTYGSLALKFKDKILDITAIKDYEVVVVGTNGKT 60  
 MK+ T LG+ AGK+A +L+K+GRSGTYG LAL DKDIL ++KDY+IVVVGTNGKT

35 Sbjct: 1 MKMKTLLGIIAGKAAQSILTKLGRSGTYGRLALACDKDILKDLSDKDYIVVVVGTNGKT 60

Query: 61 LITALTGVILKEAFGQVVTNPSGANMITGIVSTFLTAKGKSGKKIAVLEIDEASLEPRIT 120  
 LTALTGVILKEAFG+++TNPSGANMITGI STFL AKK KS ++IAVLEIDEASLEPRIT

35 Sbjct: 61 LITALTGVILKEAFGEIITNPSGANMITGITSTFLAAKNGKSKERQIAVLEIDEASLEPRIT 120

40 Query: 121 QYIKPSLFTVTNIFRDQMDRYGEIYTTQGMILDGAANAQATILANGSDPLNSKSVTNP 180  
 Y+KPSLFV+TNIFRDQMDRYGEIYTTQMI+DGA NAP+ATILANGSDP+P+SK + NP

40 Sbjct: 121 TYLKPSLFTVTNIFRDQMDRYGEIYTTQMI+VDGARNAPKATILANGSDP+P+SK+VNP 180

45 Query: 181 VQYIGFNTDKHSPRLAHNTGESILCPKCKQAILTYRLANTYANLGDYTCPCDFERPAIDYA 240  
 VQ+YGF+T KH P+LAHNTGESILCPKCK+ IL YRLANTYANLGD+ C NC F+RP LDY

45 Sbjct: 181 VQYIGFDTAKHAPQLAHNTGESILCPKCKHILQYRLANTYANLGDYVCLNQCQFQRPFLDY 240

Query: 241 LTRLTHLNTSSGPFVIDGQQYNINVGGLYNIYNALAAVVAEYFQVPSQIKDGFDKSRA 300  
 LT LT +T+ SS FVIDGQ Y INVGGLYNIYNALAAVVAEYFQV P +IK GF+KS+A

50 Sbjct: 241 LITLTAITBQSGSEFVIDGQNYKINVGGLYNIYNALAAVVAEYFQVSEKIKAGFNNKSA 300

Query: 301 VFGKQETFTTIGNKCTLVLKKNFVGASQALDNKIAFYPFSLSVLLNANYADGIDTSW 360  
 VFGKQETFTT+G+K CTL+LKNFVGASQAL+MI+LA YPFSLSVLLNANYADGIDTSW

55 Sbjct: 301 VFGKQETFTTVDGKSCITLILKKNFVGASQALEMIQLADYFSLSVLLNANYADGIDTSW 360

Query: 361 DANFETILYMNIPFIAGGVKHSIARLRVTVGYDEKRIKQADKLQDITMIEQQETIEHA 420  
 DANFE I M I EI AGGVKHSIARLRVTVG+D+ +IKQ+KLI+ I+ IE+GE +HA

60 Sbjct: 361 DANFELITQMPITENAGGVKHSIARLRVTVGFDITTKIKQARKLEQIIEETKEQEAHA 420

Query: 421 YILATYTTAMLEFREILANBNARKEM 446  
 YILATYTTAMLEFR +LA+ + + KEM

60 Sbjct: 421 YILATYTTAMLEFRSLADRHVVSKEM 446

-1522-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1388

A DNA sequence (GBSx1473) was identified in *S. agalactiae* <SEQ ID 4255> which encodes the amino acid sequence <SEQ ID 4256>. Analysis of this protein sequence reveals the following:

Possible site: 32  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3010 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:ABC84011 GB:AF080002 cobyric acid synthase CobQ [Helicobacillus  
mobilis]  
Identities = 89/250 (35%), Positives = 129/250 (51%), Gaps = 9/250 (3%)  
Query: 11 TKDYKTYLVAHLYGMLNLTNGNHLMMKYVGEKLGCOMTFDIVSLED 70  
+K TL +HLX +LNL YGD GNI+ ++ E G + + SL ++ + +  
20 Sbjct: 2 SKTNRLTLIHLYPOLLNLAYDGRNIIITLARRCEWRGITLTQVHSASLGKAAFDADLV 61  
Query: 71 FPGGGQDYBQAIIVARDLPSKKEDINKFIQNGV-VLAICGGPQLGQYIYQANGERIGHI 129  
F GGG D EQ ++ +D K G+ +L++CGG+QLLG YY GE + G+  
25 Sbjct: 62 FPGGGSDREQ/TLLPQDFQGHKGPALVEABGGLPILGVCQGYQLLGLYRTKTBEMFGL 121  
Query: 130 GVMGHYTLNQMNRVYIDIKIHDEFNE--TYGFMHQGRFTLSEDE--KPLATVYGN 185  
G+ +T +R IG++ E T GFEMH GRTEL +FL V G  
30 Sbjct: 122 GLFDWMT-EGAGSTRLIGNVVAQAPLIGEQATLNGFEMHSGRTFLSGRGGIQLAQVTAGF 180  
Query: 186 GNNKEDGTGCVHYINVPFGVYFPGKILSRNANLAVRLVATALRNKYG---KEIVLPSVEEI 242  
GNN +D EG YIN G+Y HGP+L +N LA L++ AL +YG + ++E  
35 Sbjct: 181 GNRGDDQGGGAVYKNAVGTYLHGFVLKPNPALDWLLSKALERRYGGGSLSLQDTWENR 240  
Query: 243 LSLRIPREYQ 252  
L + + +G  
Sbjct: 241 AHLGVAQRFG 250

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4257> which encodes the amino acid sequence <SEQ ID 4258>. Analysis of this protein sequence reveals the following:

Possible site: 32  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2586 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 197/260 (75%), Positives = 224/260 (85%)  
Query: 1 MTTTSLKSPPTTKDYKTYLVAHLYGMLNLTNGNHLMMKYVGEKLGCOMTFDIVSLED 60  
MTTSLKSP +DY X Y L +AHLNGNLNLTNGNHLMMKYV ERIG ++T DIVS D  
Sbjct: 1 MTTTSLKSPENQYIYDITIAHLYGMLNLTNGNHLMMKYVAEKIGARVTVDIVSIND 60  
55 Query: 61 RPDFTNYQMAFPGGGQDYBQAIIVARDLPSKKEDINKFIQNGV-VLAICGGPQLGQYIYQ 120  
P+ +Y + FPGGGQDYBQ+IIV+DLPSKK + +I NN VVLAICGGPQLGQYIYQ  
Sbjct: 61 TFRQDDYDITVFGGGQDYBQSIIVAKDLPSKKALADYIANNKVVLAICGGPQLGQYIYQ 120

-1523-

Query: 121 ANGERISGIGVMGSHYTLAQNNRNYIGDIIKHDEFNFTTYGFSNHHQRTFLSDEKPLGT 180  
 ANG +I+G+G+MGHYTLAQ+ NR+IGDIIKHDEFNFTTYGFSNHHQRTFLS DEKPLG  
 Sbjct: 121 ANGVIDGLGIMGSHYTLAQNNRNYIGDIIKHDEFNFTTYGFSNHHQRTFLSDEKPLGR 180

5 Query: 181 VTYGNQNNKEDGTGVHYKNVPGSYFHPGPILSRNNALVRLVATARNKYKGEIVLPSYE 240  
 V-YGNQNNKED TGVHYKNV+GSYFHPGPILSR NLAVRLV TAL+ KYG I LPSY+  
 Sbjct: 181 VTYGNQNNKEDGTGVHYKNVPGSYFHPGPILSRNNALVRLVTTALKKKYGSATLSPSYD 240

10 Query: 241 EILGLEIPREYGDVSKADF 260  
 +IL EI EY D+SKA F  
 Sbjct: 241 DILKQITRYADLKSASF 260

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 15 Example 1389

A DNA sequence (GBSx1474) was identified in *S.agalactiae* <SEQ ID 4259> which encodes the amino acid sequence <SEQ ID 4260>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1701(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA004402 GB:AP001509 lipocate-protein ligase [Bacillus halodurans]  
 Identities = 153/316 (48%), Positives = 212/316 (66%), Gaps = 3/316 (0%)

30 Query: 10 DPAYNVALKAYAPQKLTIDIEIFIL-WINEPATIIIGRHQNTIQKINKEFDKNGIHVVR 68  
 DP N+A+E YA + L DI+E ++L +INEP+IIIGR+QNTI+EIN E+++ NGIHVVR  
 Sbjct: 11 DFRINLAIEEYALSNL-DINETYLLFYINEPSIIIGRNQNTIEINTEYVSGNGIHVVR 69

35 Query: 69 LSGGAVYHDLNNNYTIISMNTQBSAFDFQTPSKFVIDTLAKLGVAKEFTGRNDL-EIN 127  
 LSGGAVYHD NLN++ I+ + E +FQ F+ PVI LAKGV AE GRND+ +  
 Sbjct: 70 LSGGAVYHDHGNINFSF ITKDDGSPSNFQKFDPIVIALAKLGVTAEKGRNDIAD 129

40 Query: 128 GQKFAAGNAQYKGRMMHGGCLLFDVDMVSLGQALKVSKDKIESKGIKSVRVRVNI+VDH 187  
 G+K +GNAQ KGRM HG LLFD ++ + AL VSKDKIESKGIKS+R+RV NI +  
 Sbjct: 130 GRKISGNAQFSTKGRMFSGHTLLFDSEIDHVSALNVSKDKIESKGIKSIRSRVANISEF 189

45 Query: 188 LSKKITVQBSDAILAKMEYFENDEYVLSDAELSEIQAMRDQFATWDNTYKAPET 247  
 L++KI++ +F +L ++ + +RY L+ + +EI + ++ WDNYGK+P +  
 Sbjct: 190 LTKISIDQFSLLESIFDQANTQRYKLTADDWABIEHLSKRYQNMWDNYGKSPFI 249

Query: 248 IERGVRYPAGKITTYANVNSTIKSVKINGDFGFGVKFVDIEKMLBGVRYDYKDVLAALK 307  
 ++ R+P G I V+ TI+ KIGDFRG V D+ L G+RY+ D+ AL  
 Sbjct: 250 LQHSRFPVGNIDIRLEVKQTIQCKIIGDFGFGVDVDRLEDRIVGIRYERADIBQALA 309

50 Query: 308 TVTDSQVFSRMTPREI 323  
 VD YF ++ ++I  
 Sbjct: 310 DWDVKTYGQVQEKDDI 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4261> which encodes the amino acid sequence <SEQ ID 4262>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1271(Affirmative) < succ>

-1524-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 249/328 (75%), Positives = 292/328 (88%)

Query: 1 MKYIVNTSNDPAYNVALSAYAFQRLTDIDEIFILWINEPAIIIGRHQNTIQEINKEFIDK 60  
 MKYIVN S++PA+N+ALSAYAF++L + DE+FILWINEPAIIIG+HQNTIQEINKE+ID+  
 10 Sbjct: 1 MKYIVNKSHPNPAFNIALSAYAFRELVEDELFILWINEPAIIIGKHQNTIQEINKEYIDB 60

Query: 61 NGIHVVRRLSGGGAVYHDLNLANLNTIISNNTQBGAFDFQTFSPKVIDTLAKLGVAKEFTG 120  
 +GIHVVRRLSGGGAVYHDLNLANLNTIISN T BGAFDF+TFS+PVI TLA LGV A FTG  
 10 Sbjct: 61 HGIHVVRRLSGGGAVYHDLNLANLNTIISNKTABGAFDFKTSQPVITADLGVTANFTG 120

Query: 121 RNDLEINQKFKAGNAQAYYKGRMMHHCGLLPDVMSEVLGQALKVSKDKIESKGIKSVRAR 180  
 RND+EI+G+K GNAQAYYKGRMMHHCGLLPDVM+VLG ALKVSKDKITRSKG+KSVRAR  
 10 Sbjct: 121 RNDIRIDGKICGNAQAYYKGRMMHHCGLLPDVMITVLGDALKVSKDKIESKGVKSVRAR 180

Query: 181 VTNIVDHLSDKITVQBFSDAILAQKBEYFEMDEYVLSDAELSEIQAMRDNQFATMDWTY 240  
 VTNI++ L +KITV+BFSD ILA+MRE YP+M EYVLS+ EL++I+ QF +MDWTY  
 20 Sbjct: 181 VTNILNELPKKITVEEFSDKILAKMKETIYPDMTEYVLSDBELAKIEQSAKEQFGSDMDWTY 240

Query: 241 GKAEPTYTIERGVRYEAGKITYTANVENSTIKSVKIRGDFFGVKPDDIEKMLSEVRYDYK 300  
 GKAEPTYTIER VRYEAGKI+T+ANVENS IK++KI+GDFFG+K V DIK +L G +Y+Y+  
 25 Sbjct: 241 GKAEPTYTIERNVRYEAGKISTFANVENSTIIGLKIYGDFFGKIDVQDIENLIGCKYEYR 300

Query: 301 DVLAALKTVDTSQYFSRMTPEKITKAI 328  
 DV LKT+DT+QYFSRMT EE+ KAI V  
 30 Sbjct: 301 DVFERLKTIDTTQYFSRMTVEEVAKAI 328

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1390

35 A DNA sequence (GBSx1475) was identified in *S. agalactiae* <SEQ ID 4263> which encodes the amino acid sequence <SEQ ID 4264>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.70 Transmembrane 294 - 310 ( 294 - 312)  
 40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1680 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA21748 GB:L31844 dihydroliipoamide dehydrogenase [Clostridium  
 magnum]  
 Identities = 229/589 (38%), Positives = 339/589 (56%), Gaps = 25/589 (4%)  
 50 Query: 1 MAFDVIMPKLGVDMDQGEILEWKNBQDTVNEBGVLLIEIMSDKTNMIEARDTGVLLKIV 60  
 MA V+MPKLG+ M EG ++ WRK BGD V G++L E++DK E+E+ D G++K++  
 Sbjct: 1 MAKIVMVKPLGLTMTGETLVTWKKABGQKVGKRLPEVSTDKLTNEVSESDGIVKRLK 60

Query: 61 HQAGDVVPVTEVIAYIGEEBEGVOTSSPADATITARDGQSVSGPAAPQSQTVAATPKIE 120  
 GDVV ++A IG E++ + +G S ++A +T A PKK  
 55 Sbjct: 61 VNEGDVVECTINFAIGSGADEDDISLL-----NGSSGSGSARQSITKA---FK 107

Query: 121 ELAADDEY--DIVVGGGPGAGYAAIRGAQLGKIAIVEKTEFGGTCLNVGCIPTKTYLKN 178  
 E+ A + ++VV+GGG GY AAIR AQLG K+ ++EK GGTCLNVGCIPTK L +  
 60 Sbjct: 108 EVEAVKGDNLAVVIGGPGGYAAIRAAQLGAKVLTIEKESLGTTCLNVGCIPTKTVLLHS 167

-1525-

Query: 179 ASILDLGLKVAAGRGINLASTNYAIDMDKTAFKNSVVKLTIGVGGILKANKVEIFPKLG 238  
 + + + L + K K G I + + + K V + K L G V G L L K R V + + G  
 Sbjct: 168 SGLLTMRKSGDKLGIDIGS-IVVNWKEHQKRIKIVKLVSGVSGLLFCMKVKVIGTA 226

5 Query: 239 QVNFDSKSVVIGDK-----VIKGRNVVLATGSKVSRINIFGIESPLVITSDDILDLREIPK 293  
 + + + + + N + + + ATGS I G + V + S L L P +  
 Sbjct: 227 KFBSKDTLLVTKEDGVAEKVNFDAIIATGSMFPPIEIBGNKLGSDVISTGALSBSNTE 286

10 Query: 294 SLAVNGGGVVGIELGLVWASVGVDTVTIEMADRIIPAMDKVSVLEIQILAKKGMKIKTS 353  
 S + A + + GGGV + G + H + + S G V + + IEM I + P MD + E + S + L + G + I +  
 Sbjct: 287 SIAIIGGGVVGIEFASIFNSLOCKVSIIEMLPHILFPMDREISEIAKALLRDGININNN 346

Query: 354 VGVSVIVEANNQLTKL--NKGHEVV-ADKALISGRVPPQMWOLENLEPELMEGRIRIV 410  
 V + I + + L + + GEE + + K L + + + GR + GL + + + E G I V  
 Sbjct: 347 CKVTRIBQSGDGLKVSFIDGKSGSIVDEKVLIAVGRRSNIBGLDVEIKGVKTEGGSIIV 406

15 Query: 411 NAYQETSIPCIYAPGVNTRMLAHAAVYRQVVAENALGCKRKAHLDPFPAVYTHFE 470  
 N - ET + + G I Y A G D G M A H A G V A E N + G N K K P A V Y T F E  
 Sbjct: 407 NDRHETNVEGIIYALGDCTGKINLAHVASDQGVVAERNINGQNK-KMDYITVPAVCTYKFE 465

20 Query: 471 VAMFVMTESQAREQVGDILLVGNISFTMGKRAIASNEAHGPFVKVIAEPKVKIILGVHILGP 530  
 + A VQ + TESQ + E + D VOK NG + + NE G + K + I + KY + IILGVHII + G  
 Sbjct: 466 LASVGLTESQAREKGIDYKVGKFLAANGKSLIMNTEGQVIKITDKKVEIILGVHILGP 525

25 Query: 531 AAALINEASTMINKENITLVVDAQSIHQHPTFSVMVEAFDLVLEIAIH 579  
 A + LI EA + + E T + + + H HPT E M EA L V + RIH  
 Sbjct: 526 RATDLITEAALALRIEATLEIITTTVAHPTVGERMKEAALAVNQAIH 574

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1819> which encodes the amino acid sequence <SEQ ID 1820>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.70 Transmembrane 297 - 313 ( 297 - 315)

35 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 497/591 (84%), Positives = 538/591 (90%), Gaps = 10/591 (1%)

Query: 1 MAFDVIMPKLVDMQGBEILEMKKNEDGTVNEGQVLEIMSDKTNMEIRAEEDTVLLKIV 60  
 MA + IMPKLVDMQGBEILEMKKNEDGTVNEGQVLEIMSDKTNMEIRAEEDTVLLKIV  
 Sbjct: 1 MAVEIIMPKLVDMQGBEILEMKKNEDGTVNEGQVLEIMSDKTNMEIRAEEDTVLLKIV 60

45 Query: 61 HQAGDVVPVTEVIAYIGREGEVGTSSPSA---DATTIATEDQS--VSGPAASQETVA 115  
 QAG+ VPVTEVI YIG RGE V SSP+A + T ED ++ + P AD+Q A+  
 Sbjct: 61 RQAGETVPVTEVIGYICAREGSEVSSPAASDVNVARTEDLEAGLEVPKAAQ--AAS 118

50 Query: 116 ATPKEELAADDEYDIVVGGGPRGYAAIRGAQLGGKIAIVEKTEFGGTCLNVGCIPTKTY 175  
 A PK LA DEYDI+VVGSGPRGYAAIRGAQLGGKIAIVEK+EPFGTCLNVGCIPTKTY  
 Sbjct: 119 AAPKAALADDEYDIIVVGGGPRGYAAIRGAQLGGKIAIVEKSEPGTCLNVGCIPTKTY 178

55 Query: 176 LKNAEILDLGLKVAAGRGINLASTNYAIDMDKTAFKNSVVKLTIGVGGILKANKVEIFN 235  
 LKNAEILDG+K+AGRGINLASTNY IDMDKTV FN+VVKLTIGV+GLLKANKV IFN  
 Sbjct: 179 LKNAEILDLGKIAAGRGINLASTNYTIDMDKTVDKNTVVKLTIGVGGILKANKVTIFN 238

60 Query: 236 GLGVNFDKSVVIGDKVIKGRNVVLATGSKVSRINIFGIESPLVITSDDILDLREIPKSL 295  
 GLGVNFDK+V IG + IKGRNV+IATGSKVSRINIFG+ S LVITSDDILDLREIPKSL  
 Sbjct: 239 GLGVNFDKTVTIGSQTIKGRNVVLATGSKVSRINIPGLDKSLVITSDDILDLREIPKSL 298

Query: 296 AVMGGGVVGIELGLVWASVGVDTVTIEMADRIIPAMDKVSVLEIQILAKKGMKIKTSVG 355  
 AVMGGGVVGIELGLVWASVGVDTVTIEMADRIIPAMDKVSVLEIQIL+KKGKMKIKTSVG  
 Sbjct: 299 AVMGGGVVGIELGLVWASVGVDTVTIEMADRIIPAMDKVSVLEIQILSKKGMKIKTSVG 358

- Query: 356 VSEIVEANNQLTLKIANGEVUADKALLSIGRVPQMNGLENLEPELMERGRKIVNAYCE 415  
VSEIVEANNQLTLKIANGEVUVA-KALLSIGRV QMNGLENL LEM+R RIKVN YCE  
Sbjct: 359 VSEIVEANNQLTLKIANGEVUAEKALLSIGRVSQMNGLENL--NLEMDRNRIKVNNDYCE 416
- 5 Query: 416 TSIFGIYAPGDVNGTRMLAHAARYMGEVAERNAALGN-KRKAHLDPTPAAVYTHPEVAMV 474  
TSIFGIYAPGDVNGTRMLAHAARYMGEVAERNAALGN RKA+L +TPAAVYTHPEVAMV  
Sbjct: 417 TSIFGIYAPGDVNGTRMLAHAARYMGEVAERNAALGNHIGVTRKANLYTPAAVYTHPEVAMV 476
- 10 Query: 475 GMTTEQAREQYGD+L+GKNSFTGNGRAIASNEAHGFVKVIAEIPKYKEILGVHIIQPAAE 534  
G+TEEQAREQYGD+L+GKNSFTGNGRAIASNEAHGFVKVIA+ KY EILGVHIIQPAAE  
Sbjct: 477 GLTEEQAREQYGDVLIGKNSFTGNGRAIASNEAHGFVKVIAADKYHEILGVHIIQPAAE 536
- 15 Query: 535 LINEASTIMENELTVVDVAQSIGHPTSEVMYEA7ADVLGEAHDNPPKK 585  
+INEA+TIME+ELTV ++ SIGHPTSEVMYEA7 DVLGEAHDNPPKK  
Sbjct: 537 MINEAATIMESELTVDELLLSIGHPTSEVMYEA7ADVLGEAHDNPPKK 587

SEQ ID 4264 (GBS681) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 165 (lane 2; MW 68.3kDa) and in Figure 188 (lane 10; MW 68kDa).

- 20 Purified GBS681-His is shown in Figure 240, lane 5-6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1391

- 25 A DNA sequence (GBSx1476) was identified in *S.agalactiae* <SEQ ID 4265> which encodes the amino acid sequence <SEQ ID 4266>. This protein is predicted to be dihydrolipoamide acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 46  
>>> Seems to have no N-terminal signal sequence

- 30 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.4466 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 The protein has homology with the following sequences in the GENPEPT database.

>GF1BAB04497 GB:AP001509 dihydrolipoamide S-acetyltransferase  
[Bacillus halodurans]  
Identities = 187/462 (40%), Positives = 266/462 (57%), Gaps = 26/462 (5%)

- 40 Query: 1 MAVETIMPKLGVDMMQGRILEWKKQVGDVNVNKGDLVLEIMSDKTMEIEARDSGVLLKIT 60  
MA EI MPKL MQBG +L+W K+ GD V G+ L EIM+DK N+E+EA + G LLK  
Sbjct: 1 MAKETIMPKLSTMQGTLLQWFKESGRVVEGEPLFEMTDKINIEVRYEYRGTLLKRY 60
- 45 Query: 61 HGNQDVVPVTETIGYIGAEQEVTRASSENTSVENATQVTSEPEKVEETSEPSVPAAT 120  
+G D +PV IGYIG E V +E E T E T+ P++  
Sbjct: 61 YGDDRIPVNHVIGYIGTDESVP-----TEPPGAEITASSTDEAGDHRITLVKKAPSSD 116
- 50 Query: 121 SGEKVRATPAARKLAREMSIDLALVSGTGANGRVHREDVENFKQAQPRITPLARRIAEDQ 180  
E VRATPAAR++A+E IDL V G+G GRV DV FK + TPLA+++AE +  
Sbjct: 117 R+BNVRATPAARKLAKKR+DIRQVBSGPGBRVQAVDVATFKKGGKATPLAKKVAEVK 175
- 55 Query: 181 GVDIAETTSGIRGKIVKNDVLAAMSQAAEAPVETKATPTTBEKQLPSGVEVIRMSAR 240  
GV + ++ GSG GK+ + DV A A +PVE K +K+S+R  
Sbjct: 176 GVALEKVGSGPYGKGVYREDVEHAQ-----AASPVEDKGR-----VKLSGR 218
- Query: 241 KALSQGMINSYLTAPSFTLNYDIDMTENMALRKKLIDIPIMAKTGLKVSFTDLIGAVVKT 300  
K +K M +S +AP T+ +IDM+ +R +L+ I +TG ++S+T+++ AV  
Sbjct: 219 KVAAGRMVDSAFSAPEVTTITTEIDMSITIKRSQLLGMIEQETGYRLSYTTEIVMGVAIA 278



-1527-

Query: 301 LMKPEHRYLNASLINDAQEIELANFVNIGIAGVLGDLGIVPVVHNADQMSLSDPVIASKD 360  
 LM H +NAS + EI H V+IG+AV ++ GL+VPVV + D+ L+ K  
 Sbjct: 279 LMS--HPTINDAEPFEN--EIVYEDVHIGLAVAVSGGLVVPVVKHVDKGLAQLTNECKT 334

Query: 361 VIKKTQBGKLSAEMSGSTFSTINLGMFGTKTFNFIINQNSAILGVRATPTPTTVVDGE 420  
 V ++ +L MSG TF+I+NLGM+ F P+INQP SAILGVG F +DG+  
 Sbjct: 335 VAMAARDNRLS+EMSGSTFSTINLGMFYA+DVSTFVINQPSAILGVRIGKRFVGDISQ 394

Query: 421 IVARPIAMCLTIDHRIVDGMNGAKFVLDKKNLNMENPGLLI 462  
 I RP+M L+ DHR++DG A F+ D+K++E FF LL+  
 Sbjct: 395 IELRPMTASLSDFHRVIDGAAAFITDVKSNLSQPFQLLM 436

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4268> which encodes the amino acid sequence <SEQ ID 4268>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.4774 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 354/473 (74%), Positives = 390/473 (81%), Gaps = 15/473 (3%)

Query: 1 MAVETIIMPKLGVDMQBGHILEWKKQVGDVVGSDVLEIMS DKTNMEIEAEDSGVLLKIT 60  
 MA EIIMPKLGVDMQBGHILEWKKQ GD VNEGDLIRISDKTNMEIEAEDSGVLLKIT  
 Sbjct: 1 MAPEIIMPKLGVDMQBGHILEWKKQBDTVNEGDLIRISDKTNMEIEAEDSGVLLKIT 60

30 Query: 61 HGMGDVVPVTEIGVIGABGEVIRASSENTS-----VEENATQVTESEPEKVEBTSEPS 115  
 GD VPVTE IGYIGAGEE V +SSE T+ +A + E V + P  
 Sbjct: 61 RQAGDTPVPVTEIGVIGABGEVSDT LASSEKTEIFVPASADAGAPAKENVASFA-PQ 119

35 Query: 116 VFAAT----SGEKVRATPAARKLAREMSIDLALVSGTGANGVRHREDVENFKAQAPRITP 171  
 V A +G KVRATPAARK A EM IDL V GTG GRVH+EDVENFKAQAP+ +P  
 Sbjct: 120 VAATAIFQNGSGKVRATPAARKAAENGIDLQGVGTGPKGRVHKEDVENFKAQAPKASP 179

40 Query: 172 LARRIARDQGVDAIEITGSGIRKIVKNDVLAAMSPQAAEAPVETKATPTTEK--QLPE 229  
 LAR+IA D+G+D+A ++G+G GK++K D+A + A P E KA EEK LPE  
 Sbjct: 180 LARKIAADKIDLATVSGTGNGKVKMEDIMALE---AAKPAKAPAKAEEKVDLPE 236

45 Query: 230 GVEVIKMSAMRKAKSKMTNSYLTAPSFTFLNYDIDMTENALRKLLIDPINKTGLKVSF 289  
 GVE MSAMRKAKSKMTNSYLTAFTFLNYDIDMTENALRKLLIDPINKTGLKVSF  
 Sbjct: 237 GVEHKMSAMRKAKSKMTNSYLTAFTFLNYDIDMTENALRKLLIDPINKTGLKVSF 296

50 Query: 290 TDLIGMAVVKTLMKPEHRYLNASLINDAQEIELANFVNIGIAGVLGDLGIVPVVHNADQM 349  
 TDLIGMAVVKTLMKPEH Y+NASLINDA +IRLH FVN+GIAGVLGDLGIVPVH A+H+  
 Sbjct: 297 TDLIGMAVVKTLMKPEHRYNASLINDANDIELHPFVHLGIAGVLGDLGIVPVHGNKM 356

55 Query: 350 LSLDFVIASKNVIKKTQBGKLSAEMSGSTFSTINLGMFGTKTFNFIINQNSAILGVGA 409  
 LSLDFV+ASKDVKK Q GKLK+AEMSGSTFSTINLGMFGTKTFNFIINQNSAILGVGA  
 Sbjct: 357 CLSDFVLASKNDVIKKAQKGLKAAEMSGSTFSTINLGMFGTKTFNFIINQNSAILGVGA 416

Query: 410 TIPTFTVVDGEIVARPIAMCLTIDHRIVDGMNGAKFVLDKKNLNMENPGLLI 462  
 TIPTFTVVDGEIV+RPIMAMCLTIDHR+VDGMNGAKFVLDK LNMENP LLI  
 Sbjct: 417 TIPTFTVVDGEIVSRPIAMCLTIDHRLVDGMNGAKFVLDKKNLNMENPELLI 469

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1392**

A DNA sequence (GBSx1477) was identified in *S. agalactiae* <SEQ ID 4269> which encodes the amino acid sequence <SEQ ID 4270>. This protein is predicted to be acetoin dehydrogenase (TPP-dependent) beta chain (pdhB). Analysis of this protein sequence reveals the following:

5       Possible site: 18  
       >>> Seems to have no N-terminal signal sequence

10       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1267 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9779> which encodes amino acid sequence <SEQ ID 9780> was also identified.

15       The protein has homology with the following sequences in the GENPEPT database.

      >GP:BA04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta  
           chain [Bacillus halodurans]  
       Identities = 189/319 (59%), Positives = 249/319 (77%), Gaps = 1/319 (0%)

20       Query: 11 EAINVAMSEERKDEKVFVLGELDVGVYGGDFGTSVGMLEEFQAKKRVDTPISEAAIAGSA 70  
           EAI AM+ ENRK+E VF++GED+GVYGG FG + GM+EEFG++RVR+TPISEAAI+G+A  
       Sbjct: 8 EATREAMTLENMRKNEQDFILGSDIGVYGGAGVTRGMLEEFQGSERVEVNTPISEAAISGTA 67

25       Query: 71 IGAAQTGLRPVLDLTFMDFVTIAMDIAVNQAKTINYPGQGLSTPTVTRFVAGSGSIGSRA 130  
           IGAA TG+RPI++L F DF+TIAMD +VNQ AK YM+GG P+ R +GSG G+AA  
       Sbjct: 68 IGAALIGMRPILELQFSDFTITAMDNVNVQAALRYMYGKAKVEMVLRTAGSGTGAA 127

30       Query: 131 QHSQSLEAWLTHIPGLKVAPGTVNESKALLKSSILDNHPIVFLEPKALYKKKEEVMNDP 190  
           QHSQSLEAW+THIPGLKVV P T ++K LK++I DNHPIVF E K Y K V +  
       Sbjct: 128 QHSQSLEAWMTHIPGLKVVCATAYDAKGLKAAIDNNHPIVFYEHKLCYRTKCHV-FEE 186

35       Query: 191 DFYIPLGKGDIIKRGSDTLTIVSYGRMLERVMQAAEVEAEGBINVEVDPRTLIPLOKELI 250  
           ++ IPLGK D+KR+GTD+T+V+ M+ + ++AA E+ +EGI+VEV+DPRTL+PLD+E I  
       Sbjct: 187 EYSIPLGKADVKKRGTDVTVATAMVHKALEAAVELEKEGISVEVIDPRTLVELEDTI 246

      Query: 251 IDSVKTKGLKILVNDAYKTGGFTGEIATMVAESEAFDYLDHPVRLASEDVPVPSRVLE 310  
           I SVKKT +LI+V++A K GGF GEIA++AESEAFDYLD PI RL + VP+PY+ LE  
       Sbjct: 247 IRSVKTSRLIVVHAEVKRGGPGEIASILAESAEFDYLDAPIKRLGGKVPVPIFYPTILE 306

40       Query: 311 QGILPDVAKIKDAIYKVVN 329  
           + +P V I +A+ + +N  
       Sbjct: 307 RAAIPQVPDIIIEAVKETLN 325

15       A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4271> which encodes the amino acid sequence <SEQ ID 4272>. Analysis of this protein sequence reveals the following:

50       Possible site: 18  
       >>> Seems to have no N-terminal signal sequence  
       INTEGRAL   Likelihood = -0.00   Transmembrane   81 - 97 ( 81 - 97)

      ----- Final Results -----  
           bacterial membrane --- Certainty=0.1001 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55       The protein has homology with the following sequences in the databases:

      >GP:BA04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta  
           chain [Bacillus halodurans]  
       Identities = 187/319 (58%), Positives = 244/319 (75%), Gaps = 1/319 (0%)

-1529-

Query: 11 HAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMLEEFQPKRVKITPISBAISGAA 70  
 E+ A+ AMT EMRK+E++F++G+D+GVYGG FG + GMIEFG +RV++TPISRAAISG A  
 Sbjct: 8 HAIREAMTEEMRKEDNVFLIGEDIGVYGGAGVTRGMIIEFGSERVKNTPISEAAISGTA 67

5 Query: 71 IGAATIGLRPIVDVTFMDFLTIMDAIVNNGAKNNYMGFGGLITPVTFRVASGGIGSGAA 130  
 IGAAT+TG+RP1+++ F DF+TI MD +VN AK YM-GG P+ R +GSG G+AA  
 Sbjct: 68 IGAALTGMRP1LELQFSDFITIMDNMNVQAALRYMGKAKVPMVLRTPAGSGIGSGAA 127

10 Query: 131 QHSQSLEAWLTHIPGKVVAPGNANDAGLLKSAIRDNNIVLFMEPKALYGGKBEVNDP 190  
 QHSQSLEAW+THIPG+KVV P A DANGLLK+AI DNN V+P E K Y K V ++  
 Sbjct: 128 QHSQSLEAWMTHIPGLKVVCPATAYDAGLLKAAIDNNVPIFYEHKLCIRTKCHVPEE- 186

15 Query: 193 DFIYPLGKGDIKRGSDTLTIVSYGRMLERVLQAAEEVAADGINVEVDPRTLPLDKELI 250  
 ++ IPLGK D+KR+GTD+T++ M+ + L+AA E+ +GI+VEV+DPRTL+PLD+E I  
 Sbjct: 187 ETSIPLGKADVRRKGTDTVIVATAVMVHKALAAVELEKGGISVEVIDPRTLVPLDETI 246

20 Query: 251 IESVKTKGLKMLVNDAYKTGGFGEIATMTESBAFYLDHPIVRLASEDVVPYARVLE 310  
 I SVKKT +L++V++A K GGF GRIA++I ESBAPYLD P I RL + VP+PY LE  
 Sbjct: 247 IESVKTSRLTVVHEAVRGGFGEELASIIAEEBAFYLDAPIKRLGGKVPVPIYNPTLE 306

25 Query: 311 QAILPDVEKIKAAIVGMAN 329  
 ++A+P V I A+ + N  
 Sbjct: 307 RAAPQVPDIIIEAVKETLN 325

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 286/331 (86%); Positives = 310/331 (93%)

Query: 1 MSETKVMALREAINVMASEEMRKDEKVFIMGEDVGVYGGDFGTSVGMLEEFQAKVRDTP 60  
 MSSTK+MALREA+N+AM+EMRKDE +FLMGEDVGVYGGDFGTSVGM+EEFG KRV+DTP  
 Sbjct: 1 MSETKVMALREAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMLEEFQPKRVKITP 60

30 Query: 61 ISEAAIAGSAGAAQIGLRPIVDLTFMDFVTIMDAIVNQAQKNNYMGFGGLITPVTFRV 120  
 ISEAAI+G+AIGAA TGLRPIVD+TFMDF+TI MDAIVN GAK NNYMGGLITPVTFRV  
 Sbjct: 61 ISEAAISGAAIGAAITGLRPIVDVTFMDFLTIMDAIVNQAQKNNYMGFGGLITPVTFRV 120

35 Query: 121 ASGSGISAAQHSQSLEAWLTHIPGLKVVAPGTVNESKALLKSSILDNNVPIFLEPKALY 180  
 ASGSGISAAQHSQSLEAWLTHIPG+KVVAPG N++K LLKS+I DNN V+P+EPKALY  
 Sbjct: 121 ASGSGISAAQHSQSLEAWLTHIPGLKVVAPGNANDAGLLKSAIRDNNIVLFMEPKALY 180

40 Query: 181 GKKEEVNMDPDFYIPLGKGDIKRGSDTLTIVSYGRMLERVMQAAEEVAABEGINVEVDPR 240  
 GKKEEVN DPDFYIPLGKGDIKRGSDTLTIVSYGRMLERV+QAAEEVA +GINVEVDPR  
 Sbjct: 181 GKKEEVNMDPDFYIPLGKGDIKRGSDTLTIVSYGRMLERVQAAEEVAADGINVEVDPR 240

45 Query: 241 TLIPLDKELIIDSVKTKGLKMLVNDAYKTGGFGEIATMVASBAFYLDHPIVRLASED 300  
 TLIPLDKELI+SVKTKGLK+LVNDAYKTGGF GEIATM+ ESEBAFYLDHPIVRLASED  
 Sbjct: 241 TLIPLDKELIIDSVKTKGLKMLVNDAYKTGGFGEIATMTESBAFYLDHPIVRLASED 300

Query: 301 VVPYRVLEBQGILEDVAKIKDAIKVNNKG 331  
 VVPY+RVLEB ILDFV KIK AI K+ NKG  
 Sbjct: 301 VVPYRVLEBQAILPDVEKIKAAIVGMANKG 331

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1393

55 A DNA sequence (GBSx1478) was identified in *S.galactiae* <SEQ ID 4273> which encodes the amino acid sequence <SEQ ID 4274>. Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -3.03 Transmembrane 161 - 177 ( 161 - 178)  
 60 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>

-1530-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9777> which encodes amino acid sequence <SEQ ID 9778> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04495 GB:AF001509 acetoin dehydrogenase (TPP-dependent) alpha chain [Bacillus halodurans]  
 10 Identities = 148/317 (46%), Positives = 214/317 (66%), Gaps = 1/317 (0%)  
 Query: 8 LSKSQHLDMLFKMQRIRIDVDMKHNKLVRRGFVQGMTHFSVGEAAASVGAIQDLTDSDIIF 67  
 +++++ +D+F +M IR + K ++ +G + G TH +VG+EA++VG+I L + D +  
 Sbjct: 10 MTEKGLVDLFPKQMLIRYFSEKVEFFPAKGTHIGTTHLAVGQEQASAVGSIIVLEERDKLIT 69  
 15 Query: 68 SNHRGHQQTIAKGIDIGMFAELAGKATGTSKRGSGSMHLANLEKNGYGTNGIVGGSYAL 127  
 S HRGHQ IAKG D+ M AEL G+ TG KG+GSGMH-A+++E+GN G NGIVGGG+++  
 Sbjct: 70 STHRGHGHCIAGKADVNRMMAEILFGRBTGYCKGKGGSMHIADEVGRNGLGANGIVGGGFSI 129  
 20 Query: 128 AVGAALTQQYBGTNDINIVAFSGDSATNEGSFHSVNLAJVNLFVPIFFIINNRYGISTDI 187  
 A GAALT + + +V+ F GD A+NEGSFHS+VNLA++W LEV+F NN+YG+S +  
 Sbjct: 130 ATGAALTSKMKKEGYVLFCFGDASNEGSFHEAVNLAISTWKLVPVVICENNQYGNHGSV 189  
 25 Query: 188 TYSTIKIPHYLRADAYGLPGHYVEDGNDLMAVYEMHEVINYVRSGNGPAIVEVESYRWF 247  
 I H+ RA YGPG V DGDND+ AV + ++ R G GP IVE ++YRW  
 Sbjct: 190 KEMINTEHISDRAAGTIPG-MVVDGNDVFAMVNVRRAVDARRRGGSPITVEAKTYRWM 248  
 30 Query: 248 GHSTADAGVYRIKEEVDGWRKDFVKRYRAYLIENEIATEEELAAIEAQVKEVEBGVVF 307  
 GHS +DA YRT+EE W+ KDF+ R RA L++ I TEE +I+ + ++E+ V+  
 Sbjct: 249 GHSGSDAKKRYTRBEKWKREKPIARLRATLVKEGIVTEEADSIQERAKQKIEDSVQVF 308  
 Query: 308 AEEGPFPEMSVAFEDVF 324  
 A SP F++ EDV+  
 Sbjct: 309 ARNSPFEIESLLEVDY 325

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4275> which encodes the amino acid sequence <SEQ ID 4276>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

- 40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3502 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 244/326 (74%), Positives = 278/326 (84%)  
 Query: 1 MEVRMVYLSKEQHLDMLFKMQRIRIDVDMKHNKLVRRGFVQGMTHFSVGEAAASVGAIQDL 60  
 ME MVT+SKEQHLDMLFKM+RIR+ D + NGLVRRGFVQGMTHFSVGEAA+VGA+ L  
 50 Sbjct: 1 MEAEVMVYLSKEQHLDMLFKMKEIRKFDNRINGLVRRGFVQGMTHFSVGEAAANVGAIVHL 60  
 Query: 61 TDSDIIFSNHRGHQQTIAKGIDIGMFAELAGKATGTSKRGSGSMHLANLEKNGYGTNGI 120  
 + DIIFSNHRGHQ+IAK +D+ M AELAGKATG SKRGSGSMHLA+ EKNGYGTNGI  
 55 Sbjct: 61 SYDDIIFSNHRGHQGIQAKMDLNRMMAEILAGKATGTSKRGSGSMHLADPEKNGYGTNGI 120  
 Query: 121 VGGGYALAVGAALTQQYBGTNDINIVAFSGDSATNEGSFHSVNLAJVNLFVPIFFIINN 180  
 VGGGYALAVGAALTQQY+GT+NI +AFSGD ATNEGSFHSVN+AA W LPVIFPIINN  
 60 Sbjct: 121 VGGGYALAVGAALTQQYKGINNIVAFSGDGAATNEGSFHSVNMAATWKLVPVIFPIINN 180  
 Query: 181 YGISTDITYSTIKIPHYLRADAYGLPGHYVEDGNDLMAVYEMHEVINYVRSGNGPAIVE 240  
 YGIS I +T PHLY RAARYG+PG Y EDGND+MAVYE M + + +VR GNGPAIVE  
 Sbjct: 181 YGISNNTNNTWPHLYTRAGATGVPGFCEGDNDVMAVYETMKGAVEHYRGNGGPAIVE 240

-1531-

Query: 241 VESYRWFGHSTADAGVYRTPKEEVDSWAKADPKVRYRAYLIENRIATEEELAAIEAQVIKE 300  
VESYRWFGHSTADAG YRTPKEEVD WK KDP+ +YR YL IAT++EL AI+AQV KE  
Sbjct: 241 VESYRWFGHSTADAGKYRTPKEEVDENKEDPMIKYRTRYLTSBGATDDELDAIQACVIKE 300

Query: 301 VERGVKFAEESPFDPMSVAFEDVFVD 326  
V++ +FA+ SP P++SVAFEDV+VD  
Sbjct: 301 VDDAYEFAQNSPDPELSVAFEDVWVD 326

10 A related GBS gene <SEQ ID 8797> and protein <SEQ ID 8798> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 10
MG: Discrim Score: -14.75
GVH: Signal Score (-7.5): -4.24
    Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOW program   count: 1 value: -3.03 threshold: 0.0
    INTERNAL Likelihood = -3.03   Transmembrane 161 - 177 (161 - 178)
    EXTERNAL Likelihood = 3.55   117
modified ALOW score: 1.11

*** Reasoning Step: 3

```

\*\*\* Reasoning Step: 3

```

----- Final Results -----
      bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

EF071791.298 - 1278 of 3578)
EGAD|108208|BS0806(3 - 327 of 333) acetoin dehydrogenase E1 component [Bacillus subtilis]
OMIT|NT01850951 acetoin:DCPIP oxidoreductase alpha subunit
GP|2780395|dbj|BAA24296.1|D78509 YfYK [Bacillus subtilis]
GP|2633130|emb|CB12635.1|J29108 acetoin dehydrogenase E1 component (TPP-dependent alpha subunit) [Bacillus subtilis]
GP|2957416|gb|AC05582.1|JAF006075 TPP-dependent acetoin dehydrogenase, E1 alpha-subunit [Bacillus subtilis] FIR|D69581|D69581 acetoin dehydrogenase E1 component (TPP-dependent alpha subunit) acoK - Ba
%Match = 26.3
%identity = 45.3 %similarity = 65.7
Matches = 148 Mismatches = 309 Conservative Sub.s = 67

```

[illegible]

711            741            771            801            831            861            891            921  
AVGALLTQQYSGTCNIVIAFSGDSATNEGSFHESVNLAAWNLVPVFFLNNRYGISTDITTYSTKIPHELYMRADAYGIFG  
| : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
ACGSLATACKTKOTCNVSVCFFGDGNNGCTPHEGLINLAAWNLVPVFAENNGYGREATPEPYASACSDSIADRRAAAYNTIG

[illegible]

-1532-

```

1188      1218      1248      1278      1308      1338      1368      1398
EEELAAIEAQVKEVEGEVQFAEESPPFDMGVAPEEDVFDV*NNLK*MRFTISFPYSID*KTDIRKK**AKLEKLWCAKRLM
:|: || :|: :|: | | :|: ||:| | | :|:|
--KLSDIEQRVSEBIEKAVSPSEDSPPPKDSBELLTQVYVSYKEGGM
      300      310      320      330

```

SEQ ID 8798 (GBS403) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 2; MW 64.4kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 4; MW 39.5kDa).

GBS403-GST was purified as shown in Figure 218, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1394

A DNA sequence (GBSx1479) was identified in *S. agalactiae* <SEQ ID 4277> which encodes the amino acid sequence <SEQ ID 4278>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.2464 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9775> which encodes amino acid sequence <SEQ ID 9776> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB12414 GB:Z99107 similar to ABC transporter (ATP-binding
protein) [Bacillus subtilis]

```

```

Identities = 328/643 (51%), Positives = 443/643 (68%), Gaps = 9/643 (1%)

```

```

Query: 9 MIILQGNKIERSFSGDVLFDNINIQVDQRDRIALVGRNGACKSTLLKILVGEAPTKEI 68
W+ILQ N++ +SF D + +NI ++V RDRIA+VGRNGACKSTLLKI+ G+ + KGEI
Sbjct: 1 MIILQANQLSKFSGADITLNNIKLEVNRDRRIALVGRNGACKSTLLKIAGQLSVEKEI 60

```

```

Query: 69 NKKRDLISLYLACDSRFOSENTI FQEMQLQVFDLSIREVEKRLRELELMQGVSGDLEQLN 128
K +D+++ YLAQ + S+ TI +E+L VFD L+ +EK +R +E +N +LE +N
Sbjct: 61 LKPKDITMGYLAQHTGLDLSKLTIKELLLTVFDHLKAMEKCNRAEERKNAADPFGEISIN 120

```

```

Query: 129 KTYDILSEEPREKGGFTYRSDIKAILNFKFNSDMWEMPISELSCQNTLRALAKMLLEK 188
KTYD L +EF++KGG+ YE+D++++L+G F+ + LSGQQ TRIAL K+LL +
Sbjct: 121 KTYDRLQGFQFQKGGYQYEAEDVRSVLHGLGFSHFDDSTQVQSLSGGQKRLALGKILLTQ 180

```

```

Query: 189 PELLVLDEPTNHLIDITLAWLENLYLVNYCGALLIVSHDRYFLDKVATVYDLTTHSLDRY 248
P+LL+LDEPTNHLIDIT+ WLE+YL Y GA++IVSHDRYFLDKV Y+++ +Y
Sbjct: 181 PELLVLDEPTNHLIDITLWLEHYLQYSGAILIVSHDRYFLDKVNVQVYEVRSRSECKY 240

```

```

Query: 249 VGNYSKFPDLKAEKIATEERINFEKQKRIAKLEDVQVNRIVRASTTKRAQARRKQLEKME 308
GNYS ++D KA + + K +EKQO EIAKL+DFV RN+ RASTTKRAQ+RRKQLE+H+
Sbjct: 241 HGNYSAYLQKAAQYKELKMYEKQDEIAKIQDFVDRNIRASTTKRAQSRKKQLEKMD 300

```

```

Query: 309 RLQKENVSKQSANMTFHAGKVSQGNVVLTLNAAIGYEG-VSLSEPIDLVKFKDALAIVG 367
+ KP ++KSAAN F K SGN VL +++ I YE L + + + + + A+VG
Sbjct: 301 VMSKFLGRKSNAPHDITDKQSGNEVLKVDLTISYRNQPELLTEVSPMLTRGSEALVG 360

```

```

Query: 368 PNGICKSTLILKSLVGQIPFIKGEAKLGNVETGYDDQSQSNLTKTNVLDELWDAFSTTP 427
PNGICKSTL+K+L+ + +G G+NV GYDQ Q+ LT + VLDELWD + P

```

-1533-

5  
 10  
 15

Sbjct: 361 PNGIGKSTLLKLTIDTLKPDQQTISTYGSNVSVGYDQEQAKLTSKRVLDLWDEYFGLP 420  
 Query: 428 EVEIRNRIGAFLEPSGDDVKKSVMSLGGGERARLLAKLSMENNFLILDEPTNHLIDSK 487  
 E EIR IG FLPSGDDV K V LSGGE+ARL LAKL ++ NPLIDRPTNHLID+DSK  
 Sbjct: 421 EKKEITCLGNFLPSGDDVLKFEVSLSGGGERARLAKLAKMLQANFLILDEPTNHLIDSK 480  
 Query: 488 EVLENALIEFDGTLFLVSHDRYPINRVATKVLKISDKGSTLYLGDYDYLLTKKAELELA 547  
 EVLENALI++ GTTLFLVSHDRYPINR+AT+VLE+S YLGDYDYLL KK E EL  
 Sbjct: 481 EVLENALIDYPTGLFLVSHDRYPINRATRVLELSSSHITEYLGDYDYLLTEKTEGLELE 540  
 Query: 548 RLAREEVSASKTEIDVTSD---YETQKANKQKFRKITRRVVEIRARLEVLNENNING 603  
 ++N++E KT V SD YE +K +K+ R+ RR+ EIR ++ +E + + +  
 Sbjct: 541 KXNQGE-ETDKTPATVKSDEKRSYTEEKWKKEKQRLRIETITVTQITENISRNDE 599  
 Query: 604 IMLET---NDIGHLSDLQKELSIQEBQQLLMBEWENLNRMLD 643  
 L+ + D K+ + + B + +E L+ EWE L+ D  
 Sbjct: 600 LCCDPEVYQDHEKVQAIHADNEKLNQELSLLSRWELSTEED 642

20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4279> which encodes the amino acid sequence <SEQ ID 4280>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

25

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2042 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

Identities = 473/635 (74%), Positives = 545/635 (86%), Gaps = 1/635 (0%)  
 Query: 9 MIIQQNKIERSPSGDVLFDNINIQVQDRDIALVGRNGAGKSTLLKILVGEAFTKEGI 68  
 MIIQQNK+ERSPSGDVLFI++QVD+RDRIALVG NGAGKSTLLK+LVGEE PT GE+  
 Sbjct: 1 MIIQQNKIERSPSGDVLFDNINISLQVDERDIALVGRNGAGKSTLLKILVGEETPTSGEV 60  
 Query: 69 NIKRDLGLSYLAQDSRFQSENTIPEMLQVFDGLREVERKRLRELELQMGVSGSDLEQLM 128  
 N K+DL+LSYLAQ+SRF+S+ TI++EMI+VF++LR+ EKRLR++E+ M VSG L +LM  
 Sbjct: 61 NTKKDLTSLYLAQNSRFESDQTIYEMLKVFELARQDEKRLQMEMDQATVSGQVLTPLM 120  
 Query: 129 KTYDILSEEFREKGGFTYESDKAILANGFKFNSDMWEMPISELGGGNTLRALAKMLLEK 188  
 YD+L+E FR++GGFTYESDKAILANGFKF+ MW+M I+ELSGGNTLRALAKMLLEK  
 Sbjct: 121 TDYDLATEFRFGQGGFTYESDKAILANGFKFDESMWQMTIELSGGNTLRALAKMLLEK 180  
 Query: 189 PELLVLDEPTNHLIDITIAWLENLYNYQGALIIVSHDRYFLDKVATVYVYLITSLDRY 248  
 PELLVLDEPTNHLIDITIAWLENLY NYQGALIIVSHDRYFLDKVATV DLT +LDRY  
 Sbjct: 181 PELLVLDEPTNHLIDITIAWLENLYNYQGALIIVSHDRYFLDKVATVLDLTPLNGLRY 240  
 Query: 249 VGNYSKFMDLAKKIATEEKNFEKQKKEIAKLEDFVQKNIVRASTTRAQARRKQLEKME 308  
 GNYG+FM LKAEK+ EEK F+KQKKEIAKLEDFVQ+NIIVRASTTRAQARRKQLEK+E  
 Sbjct: 241 SGNYGRFMALKAKLVAEREKQFKQKKEIAKLEDFVQKNIVRASTTRAQARRKQLEK 300  
 Query: 309 RLDKPNEVQKSNNTTHAGKVGSGNVLTLEMAAIGYEGVLSSEPIDLVKKFPAIAIVGP 368  
 RLDEP +KSA+NTPIA K SGRVVL +E AALGY LSEPI++D+ K DAIA+VGP  
 Sbjct: 301 RLDKPTGGKSAHMTPIAEKVGSGNVLTLEMAAIGYEGVLSSEPIDLVKNDLDAIAVGP 360  
 Query: 369 NGIGKSTLIKSLVGQIPFIKGSAGLGNVETGYDQSGNSLTKTMTVLDELMDAFSTTPE 428  
 NGIGKSTLIK++GQ+P +KG+ K GANVETGYDQ+QE+LT +NVL+ELM FSTTPE  
 Sbjct: 361 NGIGKSTLIKSIIGQLPLLGKLYKGNVETGYDQTSHLTSSNTVLEELWQDQSTTPE 420  
 Query: 429 VEIRNRIGAFLEPSGDDVKKSVMSLGGGERARLLAKLSMENNFLILDEPTNHLIDSK 488  
 V+IRNRIGAFLEPSGDDVKKSV+MSLGGGE+ARLLAKLSMENNFL+LDEPTNHLIDSK+  
 Sbjct: 421 VDIRNRIGAFLEPSGDDVKKSVMSLGGGERARLLAKLSMENNFLILDEPTNHLIDSK 480  
 Query: 489 VLENALIEFDGTLFLVSHDRYPINRVATKVLKISDKGSTLYLGDYDYLLTKKAELELAR 548  
 VLENALI+FDGTLFLVSHDRYPINR+ATKVLK++ GSTLYLGDYDYLL KKAELRLAR

-1534-

Sbjct: 481 VLENALIDFDGTLAFVSHDRYFIRLAKVLEITENGSTLYLGDDYLLSKKAELELAR 540

Query: 549 LNEHEVVSASKTIDVTSDYETQKANKQKFRKTRRVVIEARLEVLNDENNINGLAKLET 608  
L E E T DY+ QKANKR R++TRR ELEARLE +E I M +

Sbjct: 541 LAAGSTVETTKASAT-DYOLQKANKERRRLTRYKEIEARLSTLEKIGAIQEDMHAS 599

Query: 609 NDIGKLSDLQKLELSIQEEDOLLMEHEWENINMLRD 643  
ND +L QKE + + +EQ LNEHEW + +++

Sbjct: 600 NDIAQLLAQKQNDQLDQOEALMEHEWETABQIS 634

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1395

A DNA sequence (GBSx1480) was identified in *S. agalactiae* <SEQ ID 4281> which encodes the amino acid sequence <SEQ ID 4282>. This protein is predicted to be thiophene degradation protein F (thdF). Analysis of this protein sequence reveals the following:

Possible site: 22  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0876 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9773> which encodes amino acid sequence <SEQ ID 9774> was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4283> which encodes the amino acid sequence <SEQ ID 4284>. Analysis of this protein sequence reveals the following:

Possible site: 34  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0795 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 384/458 (83%), Positives = 427/458 (92%)

Query: 12 MSITKEFDITIAAIS\*PLGSGAIGIVRISGTDALKIATSKIYRGKDLASIOSHTLNYGHIVD 71  
P I+DEVN+ VMLAPKTFIRE+V+EINTHGSGIATVNEILQL+R GARMARPEPTK 131  
Sbjct: 1 MSITKEFDITITLSTPLGSGAIGIVRLSGTDALAIAGSVFEGKNLEQVASTITNYGHILN 60

Query: 72 PDKNEILDEVMLGVMLAPKITFREDVIEINTHGSIATVNEILQLILHKGARMARPEPTK 131  
P I+DEVN+ VMLAPKTFIRE+V+EINTHGSGIATVNEILQL+R GARMARPEPTK 145  
Sbjct: 61 PKGTIIDEVMVSVMLAPKITFREDVIEINTHGSIATVNEILQLILHKGARMARPEPTK 120

Query: 132 RAFINGRVDLTQARAVMDIRAKTDKAMDIAVKQLDGSLLKTLNNTROEILNTLAQEVN 191  
RAFINGRVDLTQARAVMDIRAKTDKAMDIAVKQLDGSLLKTLNNTROEILNTLAQEVN 150  
Sbjct: 121 RAFINGRVDLTQARAVMDIRAKTDKAMDIAVKQLDGSLLKTLNNTROEILNTLAQEVN 180

Query: 192 IDPEYDDVEEMITTLREKIQEQFQALMENLRTAKRGKILREGLSTAIIGRPNVGKSSL 251  
IDPEYDDVEEMITTLREKIQEQFQALMENLRTAKRGKILREGLSTAIIGRPNVGKSSL 215  
Sbjct: 181 IDPEYDDVEEMITTLREKIQEQFQALMENLRTAKRGKILREGLSTAIIGRPNVGKSSL 240

Query: 252 LNNLLREKAIIVTDIGTTRDVIREYVNIKGVPVKLVDTAGIRDTDDIVEKIGVERSKKA 311  
LNNLLREKAIIVTDIGTTRDVIREYVNIKGVPVKLVDTAGIRDTDDIVEKIGVERSKKA 275  
Sbjct: 241 LNNLLREKAIIVTDIGTTRDVIREYVNIKGVPVKLVDTAGIRDTDDIVEKIGVERSKKA 300



-1535-

Query: 312 LEEADLVLLVIANSEPLTLQDRSLRLSKESNRIVLIANKTDLQKIKVNEIPKVNIPISV 371  
 L+EADLVLLVIAN+SE LT QDR+LL LS++SNRI+LLANKTDL QKIE+ +LP + IPISV  
 Sbjct: 301 LQEADLVLLVIANSEKLTLDQRALLNLSDQENRIITLANKTDLQKIELQLPDYIPIPV 360

Query: 372 LENEITDKIEERINDIPFDNAGMVEHDATYLSNARIHSLIRKAVDSIAXVNRKGLKMPV 431  
 L N+NI+ IE+RIN +FFDNAG+VE DATYLSNARIHSLIRKAV SL+AVN+GL LQMPV  
 Sbjct: 361 LTNQINILIEDRIINCLFFDNAGLVEQDATYLSNARIHSLIRKAVQSLAENVGLGLKMPV 420

Query: 432 DLLQVDVTRTWELLGELITGDAAPDELITQLFSQFCLEK 469  
 DLLQVD+TRTWELLGELITGDAAPDELITQLFSQFCLEK  
 Sbjct: 421 DLLQVDLTRTWELLGELITGDAAPDELITQLFSQFCLEK 458

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1396

A DNA sequence (GBSx1481) was identified in *S.galactiae* <SEQ ID 4285> which encodes the amino acid sequence <SEQ ID 4286>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -9.18 Transmembrane 280 - 296 ( 276 - 299)  
 INTEGRAL Likelihood = -4.83 Transmembrane 249 - 265 ( 243 - 266)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4673 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD40365 GB:AF036485 hypothetical protein [Plasmid pNZ4000]  
 Identities = 88/306 (28%), Positives = 149/306 (47%), Gaps = 17/306 (5%)

Query: 1 MIVEQKFGNGFTWIN---IEABQLRTSTSEIQAKY-LDSIIITALDDYERAFMECSHIK 56  
 MI +K NG WI I AE+ T ++ +Y +D +II Y D+ E I  
 Sbjct: 1 MIKPEKTINTKWIETITQINABERAT----LEQYQIDEDIIERYTINDESINYYVD-IN 55

Query: 57 GKEVLTIIPTNTIDLKQKESYSEYVPHMTFCLSHDRILTVTRSRNSYMLALLQKYLDRNPDV 116  
 + L I L + + Y T P L L T +S + L LD NP+V  
 Sbjct: 56 EDDQLIFLAFYALDKDALRYITQPPGMLLHGKVLFTFMQSGIPEVNTALYSALD-NPEV 114

Query: 117 -SPKKFLFAAL/TLITQYFNVVSKIDREKDIARQLRQCTINKRILAMSDLETGSVYLLT 175  
 S F+ L + + + I +++ L+ L +T N L++S L+ +L+  
 Sbjct: 115 KSDVAFILETLFTVVSFIPISRAITKKNRLDQMLARKTKNSDLVLSLQLQTLFLSS 174

Query: 176 AANQIALVLQDLVHPSQRFNSVEVRKQLS---DALIRAHQLVSMQNLQVLSQLSSTF 232  
 A N L +LD P F +++++ D IE Q+ M ++ +QV+ ++ T  
 Sbjct: 175 AVQTN---LSELDLEPKTHIFGWDQDKIDLFEDVQITBQGVQRMFEIRTVVDRIHTL 231

Query: 233 NVVNNNNLNLNLTGLNIISINIAIJAATGFFGQNIPLPLETSRSSWELIVATISVLLNWI 292  
 N++ NNNLN+ + L I S+ +A+ I+GF+QNK+ LPL +W++ + SV+L V  
 Sbjct: 232 NSLANNNLNUTMKFLTINSLTMAVPTIIISFGYGMVKLEPLAQVIMVLTGSISSVLLVA 291

Query: 293 IAIILK 298  
 + +LK  
 Sbjct: 292 MLILK 297

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1536-

**Example 1397**

A DNA sequence (GBSx1482) was identified in *S. agalactiae* <SEQ ID 4287> which encodes the amino acid sequence <SEQ ID 4288>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1437 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
10 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1398**

A DNA sequence (GBSx1483) was identified in *S. agalactiae* <SEQ ID 4289> which encodes the amino acid sequence <SEQ ID 4290>. This protein is predicted to be exonuclease RexA. Analysis of this protein sequence reveals the following:

```

Possible site: 52
20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3165 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9771> which encodes amino acid sequence <SEQ ID 9772> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC12966 GB:U76424 exonuclease RexA [Lactococcus lactis]
Identities = 522/1211 (43%), Positives = 747/1211 (61%), Gaps = 73/1211 (6%)

Query: 28 KRTPEQIEAIYTFGNVLSASAGSGKTFVMVERILDKLGGVPIDSLFISTFTTKAAGE 87
35 K TPEQ EAI++ G N+LVASAGSGKTFVM +RI++K+ +G+ ID LFISTFT KAA E
Sbjct: 5 KLTPEQNFAIHSSGKNILVLSASAGSGKTFVMAQRIVEKVQGIIDRLFISTFTTKAAGE 64

Query: 88 LKERLEKKINFSLSKASRSDDLKQFLTOOLVGIGTADIGTMDAFTQKIVNQYGLTGISPI 147
40 L+ RLE+ + ++ + D+ LT L + ADIGTMD+FTQK+ + I P
Sbjct: 65 LRMRLERDLKKARQSSDDESAHRLTLALQNLNADIGTMDSFTQKLTKANFRNRVNDHN 124

Query: 148 FRILQDKNEQDVINKNEVYADLFSDYMTIGNAAS-----FIKLNVKNSGKDKSKAFREMV 202
FRIL D+ E D+I+ EV+ L Y++ + + F KL+KNFS +R+ F+++V
Sbjct: 125 FRILADQTESDLIRQEVFFQLVESYLSADESLNISKDKFEKLIKNFSKDR-NILGPKQVV 183

Query: 203 YKVYAFSQSTDNPKRMQTVFLKGAQTYIDFPAIFDQEVSSLLNVMQT--TANQLRDLTD 260
Y +Y F+ +T+NP W++ FLKG +TY +++ D +NV + T +L +
Sbjct: 184 YTTYRFA SATENPISWLNQFLKGFETY---KSLDTLSDEFTVNVKELLLTFPELLAIS 240

Query: 261 QEDYKOLTANGVPVTANYKKHKLIIENL-VHVSQDFNLLYGKKGLTINLRDITNIPSGND 319
++D+ TA LI ++ V S+D L KK + +D+
Sbjct: 241 KKDFTVTCAL-----FLSIDTDIRVSGSKDEALSAKLDKPSAQKQL----- 282

Query: 320 VTVAGVKYPIFEQAHNRIVGLGHILEVIKYGQSSLPLELLQSPVLDPSQYKELIQEN 379

```

-1537-

		V P +L + +KH ++I KYQ ++ + LQ F++DF +YL+ K EN	
	Sbjct:	--VGSKSKP--GELRKFVDKIKKQLEIKYQNAFETASDLQKFTIDPYKYTLERKKNN	338
5	Query:	AFEPSDIAHFAIQTILEENHDIRQLVQDKYHEVMVDEYQDNNTQERMLLELNGNFRPV	439
		AFE+SDIAHFAI+ILEEN DIR+ ++ Y E+M+DEYQD +HTQERMLLELNGNFRPV	
	Sbjct:	AFEPYSDIAHFAIETILEENHDIRNLREHYDEIMIDEYQTSHTQERMLLELNGNFRPV	398
	Query:	GDIKQSIYFRQADPQIENDKYKAYQNPQSGKLLILKENFRSQSEVLADSTNGVTHLMD	499
		GDIKQSIY FR ADP +F +KYK+Y + +LI LKENFRS+ EVL+ TN +F HLMD	
10	Sbjct:	GDIKQSIYGFRLADPGLFLKRYKSYDQARNPNQLIRLKENFRSQSEVLINFTNDI PKHLMD	458
	Query:	REVGDIYLDYDESHQKAS----PROQRHPRNNKTVLALDDEDDIDSDSQVDTGSPAS	555
		E+G++ Y + L G+ P +E+ + + + T E++I+D+ + IS E	
15	Sbjct:	EKLGEYTGKSEALVQGNISDYPVEAKDPYPELLLYKHWTSSEEIEDSEVK---ISDSE	515
	Query:	AKLVAKETIRLHKEENVFPDITLHNSRTNNGILQTFDQYGLPLATDGGKQNYLKSVE	615
		K A+EI +L E V +DI +LV S++ N+ I Y IP+V D G +LKS+E	
	Sbjct:	IKGAAQETIKKL-TEYGVPEKDIALIVRSKSNWNKIEDILLSYDIPVLDSEGRVDFLKSME	574
20	Query:	VNVLGLTLRSIDNPINLVALVALLRSMPGPNEDDLTRIAIDQVK+MAFYHKVLYSYHSE	674
		V++MLD LR+IDNPL D +IVA+LRSP+PGNED+LTRI+Q + +F +K +LS KE	
	Sbjct:	VLIMLDVLRAIDNPLVLDLSVAMLRSPGPNEDLTRI SVQSGRLDLPWCKILLSLAKE	634
	Query:	GHSDILITPELGSKIDHFMKTPQWDFPAKWSLYDLWIKYNDRFYDYDVGALPKAQR	734
		G + +LI L K+ F +F WR ++ L+WKIV + +Y+DVGAL ER	
25	Sbjct:	GNPELINLSLEQKLKAFNPKFTENRKLNVNKIPHRLWKIVYETTYDYDVGALKNGMR	694
	Query:	QANLYALALRANQFETGPKQLSRFIRMIDKVLNENLDLADVEVALPQNANMLTHSK	794
		QANL AL+RA +E +G+KGL +F+R+K +E NDLA V +LPONAV +MT HSK	
30	Sbjct:	QANLQALSVRAESYESSGYKGLPKFVLINKPMQNDNLASVNIKLQNAVVRMTFHSK	754
	Query:	GLEPKYVFIINIDKFSMDITSPLILSRNGIGIKTYVADMRHELE-EILPAVKVMET	853
		GLEF YVF++N+ +F+ D+ +ILSR G+G+KY+AD+ +E + P VMT	
35	Sbjct:	GLEFDYVFLANLQSRFNDRLKEDVILSRHGLGMKYIADLKAEPDVTDPFVALVMMET	814
	Query:	LPYQLNKRRLATLSEQMILLYVAMTRAEKKLYLVGKASQT---IWDHLYDLVS-ENN	909
		PY +NK +A LSE+MR+LYVA TRA+KKLYLVGK T + YD +E	
	Sbjct:	PFYVWVKDLQRAALSEBRVLYVAPTRAKKLYLVGKIKDTKAGLELYDAATLESKI	874
40	Query:	LPLASRETFVTFQDMILAVHETYYKQELFYDINPVSLEELTDHIGMNPSPFPNFK	968
		L R+ +FQ W+LA+ K L +N + +EL + + + PD K	
	Sbjct:	LSDKFRNSSRGFQHMILALQNAIK---LPMKLNVTYKDELETEKLEF+----QDFPK	926
	Query:	VNROGSEDIIVRAIS--VLESVEQINQTY--KAALPTVTRTPSQVKK-IYEPDLDIEGV	1022
		VE + D + +S +B+ +N Y +AA EL ++TPSQVKK YE L + V	
45	Sbjct:	LVHSEKFDNIMSPSDEIKKQIMNYYQPHQAATELSIQTPSQVKKRSYKQLQVGEV	986
	Query:	D-VMEITHTKTSVDVDFLPDFSTSKKQDPAALGSVAHMLQRIEMSHSVIMEDIQALTEVN	1081
		V E + ++DF DF KK A +GSA H MQ + S + Q L+E	
50	Sbjct:	QPVSEFVRVKNLDPS--DFG-PKKITAAEMGSATHSFMQYADF-SQADLPSQATLDEMG	1042
	Query:	AETSVKAAIQIEKINYPFOETSLGKYQIEVEHLHRRAFPAMLKEDPESGEKFRVGRID	1141
		+ +K I I KI F +T G+++ E V+ +EAPP+ML+ D +E++VVRGI D	
	Sbjct:	FDKIKNQIDITKILTLF-DYFEGQFSLNVDKTVKAPPSMLRTDFEAKQYIVRGICD	1101
55	Query:	GYLLENRIILFYDKTKFVNP---LELKERYOGOMALYAEALKSYEIKIDKYLILLS	1198
		G++ L ++IILEFYKTD+F N E+KERY+ QM LY+RAL+K+Y + +IDRYLILLS	
	Sbjct:	GFVKLADKIILFYDKTRFTNVGATSEIKERYKQMNLYSEALQKAYHVNQIDKYLILLS	1161
60	Query:	G-KQLLEVVKMD 1208	
		G +++ V K+D	
	Sbjct:	GPRKVFVEKID 1172	

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4291> which encodes the amino acid sequence <SEQ ID 4292>. Analysis of this protein sequence reveals the following:

Possible site: 61

-1538-

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

----- Final Results -----

5           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AA12966 GB:U76424 exonuclease RnX [Lactococcus lactis]  
 Identities = 478/1206 (39%), Positives = 700/1206 (57%), Gaps = 65/1206 (5%)  
 Query: 40 KRTAQQIEAIYTSQQNLVLSASAGSGKTFVNVERILDKILRGVSDIRLFISTFTVKAATE 99  
           K T +Q EAI++SG+NLVLSASAGSGKTFVN +RI++K+ +G+ IDRLFISTFTT KAA+E  
 Sbjct: 5 KLTPSQNSAIHSSGKNLVSASAGSGKTFVMAQRIVEKVKQIEIDRLFISTFTTKGAEE 64

15 Query: 100 LRERIEKLYSQIACTTDFQMKVYLTEQLSQCADIGIMDAFQKVSRVSGYSGISISSQ 159  
           LR R+E L +++D + LT LQ+L ADIGIMD+ F QK+ + I  
 Sbjct: 65 LRMLERLDLKKARQESSDDERAHRLTLALNLNADIGITMDSPTKIAKANFRNRVNDIPN 124

20 Query: 160 FRIMQKABQDVLRKQEVFSKLPNEFPNQKRA-----PVFRALVQVPSGNCSDTSAFRELV 214  
           FRI+ D+ E D+++QEVF+ L +++ E+ F L+KNFS ++ F+++V  
 Sbjct: 125 FRILADQTESDLIRGEVFQELVESYLSADESLNISKDKPEKLIINFSKD-RNLIQOPQVIV 183

25 Query: 215 YTCYSFSQSTENPKINLQENFLSAKTYQRLSDIPDHDIELLLLAQDTANQLRDVTDME 274  
           YT Y F+ +TENP WL+ FL +TY+ L D+ + D + + T +L + +  
 Sbjct: 184 YTYIRFASATENPISWLENQFLKGFETYKSLTDLSE-DPTVNVHVFPELLEAISIKK 242

30 Query: 275 DYGLQTKAG-SRSARYTKHLTIIEKLSNVDVDFKCLYKAGLDRILRDVIGLIPSGNDVT 333  
           D+ T S + E LS +DF +D  
 Sbjct: 243 DFTVCTALFISIDTDIRVGSSKDEALSALKDKFSA-----QKQD+V 283

35 Query: 334 VSKVKYPVFKTLHKLKQFRHLETLIMYKQDCFSLEQLQDFVLAFSEAYLAVKIQESAF 393  
           SK K + K+K H + I YQ F+ LQ F++ F+ YL K E+AF  
 Sbjct: 284 GSKSKPEGLAKFVDKIK---HGQLLEKYQNQAFELASDLQKFIIDFYKYTLERKQENAF 340

40 Query: 394 EFS DIAHFAIKILEENTDIRSQYQYHEVMVDEYQDNNHMQERLLTLLSNGHNRFPVGD 453  
           E+SDIAHFAI+ILEEN DIR++ ++HY E+M+DEYQD +H QER+L LLSNGHN FVVG  
 Sbjct: 341 EYSDIAHFAIIEERNPDIRENREHYDEIMIDEYQDTSHTQERMLLELNGHNLFPVGD 400

45 Query: 454 IKQSIYRFPQADPQIFNQKFRDYQKPEQKQKIVLLKENFRSQSEVLNVSNVPSHLMDES 513  
           IKQSIY FR ADP +F +K++ Y + +I LKENFRS+ EVLN +N +F HLMDE  
 Sbjct: 401 IKQSIYGRFLADPGLFLEKYKSYDCAENPNQLIRKENFRSQRGEVLNFTNDIFKHLMDK 460

50 Query: 514 VGDVLYDEQ+QLIAG--SHAQTVPYLDRRAQLLYNSDKDDGNAPDSSEGISSEVITVA 571  
           +G++ Y ++ L+ G S D +L+LY + + IS E+ A  
 Sbjct: 461 LGEPTYGKEBALVQGNISDYPVEAKDFPELLLYKINTSEETEDSEKVISDGEIK 520

55 Query: 572 KEIKIKHNKGVPPEDITLLVSSRTNDIISHTNPQGIPIATDGGQNYLKSVEVMVML 631  
           +EI KL + GV +DI +LV S++ N+ I Y IP+ D G+ ++LKS+EV++ML  
 Sbjct: 521 QEIKKL-IEYGVPEKDIALLVRSKNNNKIEDILSYDIPVVLDEGRVDFKSNFVLM 579

60 Query: 632 DTLRTINNPRNDYALVALLRSIPMAFDEDDLARIALQKDNELDKCLYDKIQRVIGRGA 691  
           D LR I+NP D +LVA+LRSP+F F+ED+L RI++Q +L +DKI ++ G  
 Sbjct: 580 DVLRAIDNPLDLSLVAMLRSPFGFNEEDLIRISVQGSRLR---FWDKILLSLKQSKG 636

65 Query: 692 HPLEIHETLLKGINVFLKTKLSWRRYAKLSGLYDLTWKIPNDRYFDFPVASQAKABQQA 751  
           +PELI++L KL F+ WR+ ++ L+WKI++ ++YED+V + E QK  
 Sbjct: 637 NPELINLSLQKIAKAPFKQFTBWRKLVNKPITHRLWKIYTYTETYDYVGLANGEMGA 696

70 Query: 752 NLVALALRANQPEKSGYKGLYRFIRMDKVLSTQNLADVEVATPKQAVMLTIHKSQGL 811  
           NL AL++RA +E SGYKGL++F++I+K +E NDIA V + P+ AV +MT HKSKGL  
 Sbjct: 697 NLQALSVRAESYSSRSYKGLPKFVRLINKPFGQNNDLASVNIKLQNAVVMVTMFKSKGL 756

75 Query: 812 QPFPYVILNCDKRPMSMDIHKSFIINRGHGIGIKYIADINGLLGE-TTASVSKVSMETLP 870  
           +F YVP++N RP+ D+ + IL+R+HG+GKY+AD+K T V MGT P  
 Sbjct: 757 EFDYVPLNLCGRFNRDLKEDVLRSREHGLGKYIADLKAEPVITDFPYALVQSTFP 816

-1539-

Query:	871	YQINQKELRLATLSEEMRLLYVAMTRAEKKVYFIG--ASKSKSEITDPKLL-GKLLP	926	
Sbjct:	817	Y+HK +A LSEEMRLLYVA TRA+KK+Y+GR K E+D L GKLL	876	
5	Query:	927	LALREQLLTPQDWLAIADITFSTEDLYDFVRFIEDSDLTQESVGRIQT---QLINPDL	983
		R FQ W+LA+ + L + +L E + P +L+ +		
Sbjct:	877	DKFNRSRGFQHWILALQ---NATKLPMKLNVTYKDELTEKLEPTSQDPFKKIVSESK	933	
10	Query:	984	KDNRSQSETIARALDMLSAVSQLNANY--EAAHLPTVTRPSQL-KATYEPLEPIGVII	1040
		DN S + ++EA +N Y +AA L +++TPSQ+K +YE L+ V +		
Sbjct:	934	FDINMSFSD---EIKGAQIMNYPHQATLSSITQPSQVKKRSYKQLQVSEVQV	989	
	Query:	1041	EKSSRSLSDFTLPHPSKAKVRASHIGSALHQLMQLVPLSKP--INQQTLLDALRGIDSN	1098
		+ R + + F K K+ A+ +GSA H MQ S+ + Q LD + G D		
15	Sbjct:	990	SEFVR-VKNLDPSPDFGPK-KITAEMGSATHSPMQADFSQADLPSQATLDMN-GED--	1044
	Query:	1099	BEVKTALDLKKIESFFCOTSLGQFFQTYQKHLKRAPFAILKLDPIQSSETVLRGIDAY	1158
		E+K +D+ KL + F DT GQF +EAPP+L+ D ++E+Y+RGI D +		
20	Sbjct:	1045	EKKIKQIDITTKILTLP-DTEFGQLSENVKTVKEAPFSLMRKTRFAKQIVIRGICDG	1103
	Query:	1159	FLFDDHIVLVADYTKDKYKQP---IELKKRYQQQLLYAALQTQYKLPVTKRYLWLGSG	1215
		D I+L DYKT++ E+K+RY+ Q+ LY+EAL + Y + +YL+L+GG		
Sbjct:	1104	VKLADKILIPDYKTRDFTNVSAISEIKERYKQDKNLYSEALQKAYHVNQIDKYLILGSP	1163	
25	Query:	1216	KPSIVE 1221	
		+ VE		
Sbjct:	1164	RKVFVE 1169		

An alignment of the GAS and GBS proteins is shown below.

30	Identities = 728/1211 (60%), Positives = 916/1211 (75%), Gaps = 5/1211 (0%)			
Query:	1	MMTFKFPLNPEDIAVICTEKNSSKKKQKTRPQIEIAYTFGNVNVLSASAGSGKTFVMVE	60	
Sbjct:	13	++VF PFL+PE I +Q E+ D+ QKRT +QKRYT G+N+VLSASAGSGKTFVMVE	72	
35	Query:	13	VISFAPFLSPFAIKHGLQENRCRDSQKRTAQQLIAYTSSQNILVLSASAGSGKTFVMVE	72
	Query:	120	RILDKLGRGVPIDSLPISTFTTVKAAAELEERLEKINBSLKSARSDDQLFQLTQGLGIG	120
Sbjct:	73	RILDKLGRGVIDSLPISTFTTVKAAAELEERLEKINBSLKSARSDDQLFQLTQGLGIG	132	
40	Query:	180	TADIGTIDMAPTKQVNVGGTYTGISPIRILQDKQNEGVKINEVADLFSMDYTGKNAK	180
Sbjct:	133	ADIGTMDAP KQ+V+YGY+GIS PR+QK EQDV+K+V++L+ ++N+ K A	192	
45	Query:	181	FIKLVKNSPQNRKDSKAFRMYVYKVASPQSTNPKRMQTVFLKGACQYTDPEAIPDGE	240
Sbjct:	193	FRALVKNPQNGCKDTSAPRLVYTCSPSSTENPKN+QENFLSAKTYQRLEDIPDHD	252	
50	Query:	241	VSSLLNVQCTANQLRDLTDQEDYKQLTANGVPTVANYKHLINELNVHSSQDFNLKYK	300
Sbjct:	253	IELLLNQCTANQLRDLTDQEDYKQLTANGVPTVANYKHLINELNVHSSQDFNLKYK	312	
55	Query:	301	KGLNLRADITNVISCNQDVAAGVYKPIQKHLNIRVGLKHLVYKQESLALRELL	360
Sbjct:	313	AGLRLRIDVDTYKQENQDVAAGVYKPIQKHLNIRVGLKHLVYKQESLALRELL	372	
60	Query:	361	QSFVLDPSEQYLQKICGNVDSGVYKIAVFAKILIEHNDIRQLYQDKYHEWVDKEYQDN	420
Sbjct:	373	QDFVLDPSEAYLAKIQSSAFDSIAHFAIKILEENDIRQSYOCHYHEWVDKEYQDN	432	
65	Query:	421	HTQHEVLDLNSGNHPRMVDGDKQSTYFRACQDPPIKPNKYQKAYQCNPSQKLLIKGNF	480
Sbjct:	433	HMQRELLTILSNGHPRMVDGDKQSTYFRACQDPPIKPNKYQKAYQCNPSQKLLIKGNF	492	
	Query:	481	RQSGEVLSTNVSTHLMDEYGVGDIYDSESHQLGASPRQERHPRNKTQVLLD+DED	540
Sbjct:	493	RQSGEVLSTNVSTHLMDEYGVGDIYDSESHQLGASPRQERHPRNKTQVLLD+DED	552	

An alignment of the GAS and GBS proteins is shown below.

30	Identities = 728/1211 (60%), Positives = 916/1211 (75%), Gaps = 5/1211 (0%)			
	Query: 1	MCTFPKFLNPEDIAVIQTEKNSDKKQKRTPEQIEAIYTPNNVLVSAGSGSKTFVMVE	60	
		+++F PFL+PE I +Q E+ D+ QKRT +QIEAIYT G N+LVSAGSGSKTFVMVE		
35	Sbjct: 13	VISFAPFLSPKAIKHLQENRCDQKQKRTQIEAIYTPNSQNLVSAGSGSKTFVMVE	72	
	Query: 61	RILDKILRGVPIDSLFISTFTFVIAAGELIKERLEKKINESLSAESDLEQLFQLQGLVGI	120	
		RILDK+LRGV ID LFISTFTFVIAA EL+ER+E K+ + +K +L+QL +		
	Sbjct: 73	RILDKILRGVSDRLFISTFTFVIAATELRERLENKLYSQAQTIDPQMKYLTBGLQSLC	132	
40	Query: 121	TADIGTMDAFTQKIVNQYGYTLGISPIFRILQDKNQGQDVIGNEVYADLFSDYMTGKNAS	180	
		ADIGTMDAF QK+V++YG+G+IS FRI+QDK EQDV+K EV++ LF++M K A		
	Sbjct: 133	QADIGTMDAFQAQKVSRYGYSIGSSQFRIMQDAEQDLVKGQVPSKLNFSPNQKQAPV	192	
45	Query: 181	FIKLVKNPFGNKRKDSKAFREMYKVA?SQSDNPKRMQTVFLKGAQYTDPEAIPDQE	240	
		F LVKNPFGN KD+ AFRE+VY Y+FSQST+HFK W+Q FL A+TY E IPD +		
	Sbjct: 193	FRALVKNPFGNCKDTSAFRELVYTCYSFSQSTENPKILQENFLSAKYTORLEIDPHD	252	
	Query: 241	VSSLNVMQ?TANQLRDLTDQEDYKLTAKGVPTANYKHILKILRNLVHNSQDFNLLYK	300	
		+ LL MQ ?ANQLRD+TD EY QLT G +A Y KHL IIR L W+DF LYK		
50	Sbjct: 253	IELLLANQ?TANQLRDVTDMEYQGLTKGSRSAKYTHLITITKLSDWVRDPKGLK	312	
	Query: 301	KGLVNLARDITNVIPSGNDVTVAGVKYPIPKQLHNRIVGLKHLIVIPKYGQSRSLLELL	360	
		GL L RD+T +IPSGNDVTV+ VKYP+FK LH ++ +HLE I YQ + LLE L		
	Sbjct: 313	AGLDRILRDVTVGLIPSGNDVTVSGVKYVPVKTLHAKLQKPRHLETTILMQCKCFLEQL	372	
55	Query: 361	QSFLDPSQYQLQKICQNAPEPSDIAHFAIQILEENHDIRQLYQDKYHEVMVDXYQNN	420	
		Q FVL FSE YL KIQS+AFPEPSDIAHFAI+ILEEN DIRQ YQ YHEVMVDXYQNN		
	Sbjct: 373	QDFVLAFSEAYLAVKIQS+AFPEPSDIAHFAIKILEENTDIRQSYQYHEVMVDXYQNN	432	
60	Query: 421	HTQERMLELLSNGHNRPMVGDIKQSIYRPRQADPQIPNDKYQAYQNPSSQGLKILKENF	480	
		H QER+L LLSNGHNRPMVGDIKQSIYRPRQADPQIPN K++ YQ P QQR+I+LKENF		
	Sbjct: 433	HMQRILLTLLSNGHNRPMVGDIKQSIYRPRQADPQIPNQKFRQYQKPKQKGVLLKRN	492	
	Query: 481	RSQSEVLSDTNSVPTHIMDEEVGDILYDESHQKAGSPRQQRHNNKTVLLLLTIDED	540	
		RSQSEVL+ +N+VF+HIMDE VGD+LYDE HOL AGS Q + + +Q+LL ++D+D		
65	Sbjct: 493	RSQSEVLVSNVAFSHIMDESVGVLYDEQHQLTAGSHAQIVPYLLDRAQLLLYNSDKD	552	

-1540-

Query: 541 IDSDSQQYDISPRAEKIVAKEIIRLHKEENVPPQDITLLVSSRTENDGILCTDFPQYGP 600  
 ++ S IS +E +VAKETI+LH ++ VFP+DITLLVSSRTEND I TP++YGIP  
 Sbjct: 553 -GNAPSDSBSGISPSEVTIVAKSIIKLHNDGVPFDITLLVSSRTENDIISHTFQLQYGP 611

5 Query: 601 LVIDGGGQNYLKSVEVVMVLMOTLRISINPLNDYALVALLRSPMGPNEDDLTRIAIQ-- 658  
 + TDGG+QNYLKSVEVVMVLMOTLR+I+NF NDYALVALLRSPMF F+EDDL RIA+Q  
 Sbjct: 612 IATDGGQQNYLKSVEVVMVLMOTLRITINNPNNDYALVALLRSPMFAPFEDDLARIALQKN 671

10 Query: 659 --VMAFYHKVKLSYHKBGHHSDLTIPELSSKIDHFMCTQPTWEDFANWHSLVDYLIWKY 716  
 K Y K++ + G H+LI L K++F+K++WE +AK SLYDYLWKL+  
 Sbjct: 672 ELKDKCLYDKIQRAVIGRGAPHELHUTLGLKLVFLKTLKSWERTAKLGLSLYDYLWKL 731

Query: 717 NDRFYDVTGALPKABQRQANLYALALRANQPEKTGFKELSRFIRMIDKVLNENNDLADV 776  
 NDRFY+D+V + KAPQ QANLYALALRANQPEK+G+KGL RFI+MIDKVL +NLDADV  
 Sbjct: 732 NDRFYDFVASQAKABQAQANLYALALRANQPEKSGYKGLYRFIMIDKVLNENNDLADV 791

15 Query: 777 EVALPQNAVNLMTIHKSGLFEKCYVFILNIDKKPSMVDITSPLLSRNQGIGIKYVADMR 836  
 EVA P+ AVNLMTIHKSGL+F YVFILN DK+FSM DI IL+R GIGIKY+AD++  
 Sbjct: 792 EVATPFQAVNLMTIHKSGLQPPYVFILNCDKRFMSMDIHKSFILNRQHGIGIKYLADI 851

20 Query: 837 HELSEELIPAVKVSMTLPYQLANKRELRLATLSQMRLLYVAMTRAENKLYLVGKASQTK 896  
 L E L +VKVSMETLPYQLANK+ELRLATLSR+MRLLYVAMTRAENK+Y +GKAS++K  
 Sbjct: 852 GLLGETTLNVSQVSMETLPYQLANKQLRLATLSRMRLLYVAMTRAENKVFYFGKASQK 911

25 Query: 897 WADHYDLVSENNHLPASRETFVTPQWLLAVHSTYKQELFYDINFVSLEETDHHIGH 956  
 + D LPIA RE +TPQWLLA++ + +L++D+ F+ +LT +G  
 Sbjct: 912 SQBITDPKGLKLLPLALREQLLTPQWLLAADIPTSEDTLYDFVRFIESDLTQBSVGR 971

30 Query: 957 VNPSPFPNDKVENRQSEDIIVRAISVLESVEQINOTYKAAIELPFTVTPSQVKIYEP 1016  
 + NPD+ +NRQSE I RA+ +LE+V Q+N Y+AAI LPTVTPSQ+K YEP+  
 Sbjct: 972 LQTPQLLNPDLLKNNRQSEDIIVRAISVLESVEQINOTYKAAIELPFTVTPSQKAYEPL 1031

Query: 1017 LDIBGVDMETITKTSVDKLPDPSTSKQDPAALGSVHLMQREMSHVVNEDIQKA 1076  
 L +VD++E +++ DF LP FS K++ +SSR+H+LMQ +S + + +A  
 Sbjct: 1032 LEPGVDTIIEKSSRSLSDFLPHFSKKAKVEASHIGSALHQLMVLPKSPFNQQTLLDA 1091

35 Query: 1077 LTVENASTVGAIAIQIKKINYPQETSLQKIQEVEHILREAPFAMLKEDPESGEKPVV 1136  
 L +++ VK A+ ++KI FF +TSLG++ Q +HL+REAPFA+LK DP S E++V+  
 Sbjct: 1092 LRGIDSNREVKTALDKKIESFPDCTSLGQFPQTYQWLYREAPFAILKIDPISQEEVVL 1151

40 Query: 1137 RGIIDGYLLENRIILFDYKIDKFNPLELKERVYQGMALYAEALKSYEIKIDKYLIL 1196  
 RGIID Y L++ I+L DYKIDK+ P+ELK+RYQ Q+ LYAEAL ++Y++ +YL+L  
 Sbjct: 1152 RGIIDAYFLPDHIVLDYKIDKFKQPIELKRYQCCLELYAEALQTYVLEPFRKYIVL 1211

45 Query: 1197 LGGKQLEVVM 1207  
 +GG + E+V++  
 Sbjct: 1212 MGGKPEIVEV 1222

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1399

A DNA sequence (GBSx1484) was identified in *S. galactiae* <SEQ ID 4293> which encodes the amino acid sequence <SEQ ID 4294>. This protein is predicted to be exonuclease RexB. Analysis of this protein sequence reveals the following:

55 Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0660 (Affixative) < succ>  
 60 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1541-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AC12965 GB:U76424 exonuclease RexB [Lactococcus lactis]  
 Identities = 363/1093 (33%), Positives = 604/1093 (55%), Gaps = 67/1093 (6%)

5 Query: 1 MKLLYTDINHDTEILVQAHAHMAAGWRIFXIAPNSLSFKEKRAVLEHLQ---EASFA 57  
 M++LYT+I D+TE L+ A E +++YI P++SFEKE++LE L+ A F  
 Sbjct: 1 MEILYETITQDLTGELIALSELEKKNRKYVYIVPSMSFKEKLELERLAGKSDTAVFD 60

10 Query: 58 ITITRFAQLARYFTLMQP-NQKESLNDIGLMMIFPALASPEDGQLKVGRLQDASFTS 116  
 ++TRF QL YF ++ K L +GL+M+R L SF+ ++ ++ L+ L+ A+  
 Sbjct: 61 LLVTRFKQLFYFDFRREKATMKELGTGLSMLERKVLRSFKKDLPIFYLSIQDAGFLE 120

15 Query: 117 QLVDLYKQLQANLSLEKLYLHSPKFFDLAIFLVVSDLLREGEYDNGQSKIAFFTEQV 176  
 L+ L EL TANLS+ L ++ +LA F + EY N S+ FT +  
 Sbjct: 121 MLIQRAELLTANLSVENLDPNPKNQELKKILAKFEARLSV---EYANSEPGDPTNRL 176

20 Query: 177 RSGQLVDVLKNTILVDGPTFRPSAKERALLKSLSSRCQSIILGAYASQKAYKANFTNGNI 236  
 G+ D LK+ +I+DG+TRFSAREE I+S+ + ++G Y+ + A + I  
 Sbjct: 177 VDGFDQLKQDVITIIIDGYTRFSAREELLFESIQEKVARFVVGTSDENSITAG--SETI 234

25 Query: 237 YSAGVDFLRYLATTPQTKPEFPLSKWESKSGFEMISK-----NIEGKDFNESHLLDIT 291  
 Y + T F+ K L K S+ E+ SK ++++ T+ L+ A+  
 Sbjct: 235 YVGTSQMI---TRFRMKFPVELKRIASSAVNEVYSKLTIRILDLSRPFYTDKEIKLQAE 290

30 Query: 292 AKDCITIVBCINQKDEHVARAIRKLYQCHRYKDILVLLQDVSQYKLSKIFPCYDI 351  
 + IWE NQK E+E VA+ IRQK+ QG +KD VL+GD +Y++ L ++E Y+I  
 Sbjct: 291 DEKYFRWAEANQKVEIRVAREIRKQLIQGAFFKDPVTLVGDPAAYETITKLVDYAEI 350

35 Query: 352 PYYGKAEHMAHFLVHMFESLSRIKKRYFRADVLNPKTOIVGESSQD--LDYFEAY 409  
 P+++ + E+M+ HPLV F +SL IK+ +R +DV+NL K+ +Y + + D+ +DYFE Y  
 Sbjct: 351 PFFYAGESMSQHPLVIFPESLEPAIKQNNYRTDDVNNLLSKSVYTDANLQDEVIDYFHY 410

40 Query: 410 ISYADIKGPKKFFTPFVGAQKFDLGRNATIRQSIL---TLESFV-KTKQGGIKTFLAQ 469  
 + I G KKF +F+ ++ + +N+R+ LL +PL+ F+ +K+ G K ++  
 Sbjct: 411 VQYKISGRKKFTSEFTE-SEFSQIELVNMREKLLGSESPLQVFLGNKKIKGKKGVSD 465

45 Query: 466 FMFFLTQVGLSINLSRLVQMS-ENVEQE---KHQEVNKTFTDILEQPTFGCEKRLALDE 521  
 L + N++ +HE + KH++VW+ L F+ F EKL E  
 Sbjct: 470 LQGLLENQNVNMMNAYFSAELQNEHQADKHQVWQMLISTNEFLVPSDEKLKQVE 521

50 Query: 522 FLSLINSQNMCAEYRMVPATVDVVTVKSVDLVEPHSNCFVYALQMTQSHFPAQNKSLI 589  
 FL +L +G+ A+YR +PA VDVV VK Y+LVEP +N++YA+G++Q++FP+I +N +L+  
 Sbjct: 530 FLDILLAGLNAKRYQIPANVDVNVKDYELVEPKNKYIYAIGLSQTNPRKIQSITLL 589

55 Query: 582 SDIERQLINDANDTDGHFDIMTQENLKNHFAALSLNAAKQELVITTPOLLNESQDMS 641  
 SD ER IN D+ ++ N+KN F LSL N+AK+ LVL++PQ++ + S  
 Sbjct: 590 SDEERLEINQTTDENQFIEQLNVANYCKNQPTVL+SLNSAKESVLNMQIMANQEQSES 649

60 Query: 642 P-YLVELRDIGVPPNIGKR-QSLKEEADNIGNYKALLSRVVDLYRSALKDKMTKEE-QTF 696  
 P + + L+D K + +LE + +IGN +++++ + R + + E T E+ + F  
 Sbjct: 650 FVFQLFLKDADEKILQKIQGNLPSLKHGHSRSVIAMGQIKRSLVSEHSETSDEKRVF 709

65 Query: 699 WSVAVRYLRQQLTSKGIEHPIITDGLDTVTYSSDWIRFRFDDPLKLSGALTTFYNG 758  
 WS R L+ + + + +DTV + + D + + + D + S S+ FYN +  
 Sbjct: 710 WSSIFRILVKSNAFPQKILLDLAKDITVNLAPDTLEQIY--GDKIYASVSSPERFYNCE 767

70 Query: 759 YKYFLQYVLGHRQDSIHPIMRHGTYLHRVFEILMKQOI--ESPEEKIASDKNTQGE 816  
 Y+YFL+ L LE ++I + + G + H VFR +MK + E+F+EKL + + + +  
 Sbjct: 768 YQYFLENTLSLETPEINIDINSKIVGNPFHEVPEKVMKRTDLASNNDEKLLTLVLQEDVKN 827

75 Query: 817 DVPKLSYSEDAESRYSLEILADIANATATILR-----ERRPELM 865  
 + +++DA +R++ LE+I R TAT+L+ D T+ +E E  
 Sbjct: 828 --YSRYFTQDARFTWSNLESEIVRQTATVLAIVATVDELKTLATSSSPGLPKSELGNPS 885

80 Query: 866 IDNTIKINGIIRIDRLSDGLGVVDYKSSAQKDFIQKPYNGLSPLQVITYALSDEKVE 925  
 +D+ I + G IDR+D+LS IG +DYKSSA F +Q+ Y+GLS Q +TY+D I K+  
 Sbjct: 886 VDD-IYLRGRIRIDQLSDYLGADYKSSASFKLQRAYDGLSLQPMYTLQVI---KQA 941

-1542-

Query: 926 BQKPPITGAMYLHMQEPKQDLGSKIKNDDLVTKNHQALTYKGLFSEARKEFLANGKXHL- 984  
I+GA+YL + +LE+I L ++ +++ Y+GL E E+ G ++  
Sbjct: 942 FPAQKIQGALYLPQKQPNINLSRINQISEIANILKESNRYBGLVLEDAARQI-KGIGNIA 1000

Query: 985 --KDSLYSRTEIAILQAHNQSLYKKASETIKSGKFLINPYTEDAKTVDDG-----Q 1033  
K +++E E L N+ Y+ A ++K GK INP +++ +D  
Sbjct: 1001 LKKTINITYNEEFQGLLKLNERHYRAAGQRLKGGKIAINPIMKRSRGIDQSGNVGRGCRYCP 1060

Query: 1034 FKSITGFEADRH 1046  
KSI FEA+ HM  
Sbjct: 1061 LKSIICRPEANIH 1073

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4295> which encodes the amino acid sequence <SEQ ID 4296>. Analysis of this protein sequence reveals the following:

Possible site: 23  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1891 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 546/1075 (50%), Positives = 758/1075 (69%), Gaps = 11/1075 (1%)

Query: 1 MKLLYTIDINHMTTEILVNQAARAAGMRFIFYIAPNSLSFKERKAVLENLPQESFAIT 60  
MKL+YT++++ MTEILVN+ A A+ G+R+FYIAPNSLSFKER VL LP+ +F+I +  
Sbjct: 1 MKLIYTR+GYSMTTEILVNEARKAAGQYRVFYIAPNSLSFKERKAVLENLPQESFAIT 60

Query: 61 TRFAQLARYPTLANQPMKESIMDIQLAMIFYRALASFDGQQLK+VGRLLKQDASPSISLV 120  
TRF Q++RYPT+ K+ L+D LAMIFYRAL + L +GRK+ ++ FT QLV+  
Sbjct: 61 TRFVQMSRYTTFVSSPSKQHLDDTLTAMIFYRALMQLKPEDLPSYGRLLQNSVPTQLV 120

Query: 121 LYKELQTANLSILELKYLSPEKPEDLLAIFLVVSDLLRGEYDNGSKIAFFTEQVRSQ 180  
LYKEL+ A L\$+ +L L P+K BDL+ I + ++ ++Y+ S + F ++G  
Sbjct: 121 LYKELKNAQLSVHDLTGLDHQKQEDLKIIELAETIMIQDQVYNDSPSQPARAILGL 180

Query: 181 LDVLDLNTLIVDGFTRFSABEELIKLSLRQCEIIGAYASQKAYKANFTNGNIYSAG 240  
L+ L T++++DGF+RFSABE+ L+ L++ QGR+IIG+Y SQAY+ +F F+K+  
Sbjct: 181 LNNQLSKTVTVIDGFSRFSAREDYLLGLNANNQGEVIIGSYVSQKAYKQKSPFKNGIYAS 240

Query: 241 VDFRLVLAATTFQTKPEFILSKWBSKSGFEMISIMIBGKHDFNTNSHILDDTAKDCITIE 300  
+ FL+ LA + KP F S K F +++ E HDE+ L + D +++  
Sbjct: 241 LHFLQDLAQKHILKFPVATSNQVFPAPSKRLQLFQATHDPSQVWQLQKSDLDHPSLW 300

Query: 301 CINCKDEVHVARAIRKLYGGRYKXDLVLVLGVDVSYKLQSLKIPQYDIPYFGKAET 360  
C +QK+E+EHVA++IRKLY+GRYKXDLVLGLD+D+Y+LQ+ IP++++IPY KAE  
Sbjct: 301 CHHQKEIEHVARSIIRKLYGGRYKXDLVLGLDMDATQLQGIPIPKPEIPFYLGKAE 360

Query: 361 MAANPLVHFMDSLRIRKRYPRADVNLKFTGTYGRISQDLDVFRAYISYADIKGPK 420  
MAHPLV F++EL R +RY +R EDH+L K+G++G D+D FE Y +ADIK G  
Sbjct: 361 MAHPLVQPIEELERSQRYNRREDILNMLKSLGFCGPDSDIDRFRYTFQADIKGPTK 420

Query: 421 FFTDFVW-GAKKFDLGRLLNTIRQSLITFLFSVFKTKKQGIKTNLQPMFFLTQVGLDNL 479  
F F + ++++ L LN +RQ ++ EL+ K++KQ G +++ FL ++ L++  
Sbjct: 421 PSKFPFTINSSQYPLDPLNEMRQDVLPLQELFKSQKQLGASLVKLLILFLKRLKRLNM 480

Query: 480 SRLVGQMSNEKSEKQKQEVWKTFTDILBQQTIPQGEKLNLDLFLSLNSGMQAEYRNP 539  
L S+ E EK++EVWK FTDL F IPQGEKL L + L+L+ +GM A+IR+VP  
Sbjct: 481 QSLA--QSQLEVRKNEVWKTFTDILSRHRI PQGEKRLSLDCLALIKYKMSAQYRVVP 538

Query: 540 ATVDVVTYKSYDLVEFHSNPFYALQMTQSHPPKIAQNGSLSIDIERGLINDANDTIGHT 599  
AT+DVVT+KSYDLV+PES FVYA+G+TQSHPPK + L+SD ER IN+ + HF  
Sbjct: 539 ATLDVVTYKSYDLVQFHSKPPFVYALGITQSHPPKQIHSGSLSDQERARINEIRNY-RHF 597



-1543-

Query: 600 DIMTQENLKKHFAALSLFNAAKQELVLTIPQLLNESDQMSPYLVLRDIGVPFNHGR 659  
 DI + RN KKH ALSLFNAA +ELVLT+ +NE+ D +SPYL EL + G+P KG+  
 5 Sbjct: 598 DIASAENSKKHQHTALSLFNAATKELVLSVSTVINSTFDOLSPYLKELINGGLFLIDKKG 657

Query: 660 QSLKEADNIGNYKALLSRVVDLYESAIDKEMTKKEQTFFMSVAVRYLRQLTSKGIETPI 719  
 L + +IGNYKALLS+++ + R + EM+ +++ FW+V +RYLR+QL + +E+P  
 10 Sbjct: 658 NYLSYNSDIGNYKALLSQITAINRQL- IEMSDQDMFWTVVLYRLKQLRKQQLLPT 716

Query: 720 ITDSLDTVIVSSDWMTRRFPEDDPLKSSAATTFVYNQYKVPFQVVLGLESQSIHPDM 779  
 L T +S +V+ FP+ PLKLS+ALT FVYNQY YFL+VVL L + +SIHPD  
 15 Sbjct: 717 SDYRLSTKPLSKVIEVCFPKGIPILKLSATAITVFYNNQYNYFLKVLINLNTESIHPDS 776

Query: 780 RHHTYLYHRVFEILIMQOGIESPEEKLSAINKINQDRDVKLSYSDAESYSLLELSDI 839  
 R HQ YLHRVFE LMK+ E F+ KL AI TNQR F+ +Y +AE+ YSL ILWDI  
 20 Sbjct: 777 RINGQYLRVFRILMKDHTQEPFDNKLKQAIYHTNGESFFQOVQDNAAEYSILAIEDI 836

Query: 840 ARATATILRQDSQMTVESEERFELMINDTIKINGIIDRIDLSGSLGVVDYKSSAQKF 899  
 R+TA IL+ + + V +E+ F+L + N I +GIIIDRID+LSDGSLG+VDYKSSA +F  
 25 Sbjct: 837 VRSTAPILQANQIQVIDQEFNQFLDMGNEILVHSGIIDRIDQLSGSLGIVDYKSSANQF 896

Query: 900 DIQKFYKLSPOLVTYIDALSR--DKEVEQKPPFGAMYLHMQEPQDLSKIKNLD-LV 956  
 DI FYKLSPOL+TY+ A+ + ++ Q +FGAMYLH+Q+P+ DL K +D+ LV  
 30 Sbjct: 897 DIGTFYKLSPOLMTYLAALKQIAPHDINQ---LFGAMYLHQLDQKLDVTFKQIDNTLV 953

Query: 957 TRHQHAIYKGLPSAEKRFPLANGKTHLDSLYSETIAILOAHNQSYLYKASETISKG 1016  
 ++ALT+YKG+PSE EKE L+ G Y K+LYS E+ L +N+ LY KA++ IK G  
 35 Sbjct: 954 BSIYKALTYKGIPESEVKEHLSTGAYCTGNALYSNDELETLLNKNYLYLNAKHHKKG 1013

Query: 1017 FLINPYTDAKTVDGQDKFSITGFADRHMARALYKLPAAKROGQFLTMQOE 1071  
 FLINPYT D KTV GDQ K+IT FEAD M +AR L LPAKEK+ FLITLM+AE  
 40 Sbjct: 1014 FLINPYTSDGKTVDGQDLKAITFEADLDMQARRLLTLPAAKEKCEFLTARKE 1068

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1400

A DNA sequence (GBSx1485) was identified in *S.agalactiae* <SEQ ID 4297> which encodes the amino acid sequence <SEQ ID 4298>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.80 Transmembrane 51 - 67 ( 44 - 69)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4121(Affirmative) < succ>  
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8799> which encodes amino acid sequence <SEQ ID 8800> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10  
 McG: Discrim Score: -20.62  
 GvH: Signal Score (-7.5): -6.25  
 Possible site: 31  
 >>> Seems to have no N-terminal signal sequence  
 55 ALOM program count: 1 value: -7.80 threshold: 0.0  
 INTEGRAL Likelihood = -7.80 Transmembrane 47 - 63 ( 40 - 65)  
 PERIPHERAL Likelihood = 3.34 26  
 modified ALOM score: 2.06

60 \*\*\* Reasoning Step: 3

-1544-

----- Final Results -----

bacterial membrane --- Certainty=0.4121 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:AACT75528 GB:AB000334 orf, hypothetical protein [Escherichia coli K12]  
 Identities = 138/297 (46%), Positives = 193/297 (64%), Gaps = 16/297 (5%)

10

Query: 5 MKIDDLKRSNVDERRSSSGSFSGQSGLPILQLLLRGSWTKLVVLIILLLG--GG 62  
 M+ R+SDNVEDRR+SSG S GG G + S K L++LI++L+ G G  
 Sbjct: 1 MRWQRRRESNVDERRSSSGGF-SMGPGFRL-----PSGKGLILLIVLVVAGVYV 52

15

Query: 63 GLTSINDFSPPSYQSQNVSRVDNSATRQIDFVNKVLSTEDPWSQEPQTQSGNYK 122  
 LT + +++S + D +A F + +L +TSD W Q+P+ G Y+  
 Sbjct: 53 DLTLGLMTGQFVSQQQSTRSISFNDEAAK-----FTSVILATTEDTWGQPFKMG-KTYQ 106

20

Query: 123 EPKLVLYTNSIQTGIGESAGPFYCSADKKIYLDISFYNELSHKYATGDFAMAYVIA 182  
 +PKLV+Y +TSGG G+S GPFYC AD +Y+D+SFY+++ K GA GDFA YVIA  
 Sbjct: 107 QPKLVMYRGWTRTGGAGGSGINGPFYCPADGTIVYIDLSFYDMKDKLGADGDFAGYVIA 166

25

Query: 183 HEVGHGHIQTELGIMDKYMRHGLTKKRNALNVRLELDADYAGVWYHRYGRNLLGQ 242  
 HEVGH+Q LGI K +++ T+ E N L+VR+ELQD +AGVW H ++ +LE G  
 Sbjct: 167 HEVGHVHQLLGLIEPKVRQLQCNATQAEVNRLSVRMELOADCFAGVWGHSMQQQGLVTE 226

Query: 243 DFEENVAUAAVAGDDTLQKETYGKLVDPSTHGTAEQRQRWPKGPFQYQIDGHQDTF 299  
 D EEA+VAA A+GDD LQ+++ G+VDPSTHGT++QR WF GF GD +TF  
 Sbjct: 227 LEEALNAAQAIGDRLQQCSQGRVVPDSTHGTISQQRYSWFKRGFDSGDPACQNTF 283

30

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4299> which encodes the amino acid sequence <SEQ ID 4300>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -6.42 . Transmembrane 48 - 64 ( 41 - 67)

35

----- Final Results -----

bacterial membrane --- Certainty=0.3563 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40

The protein has homology with the following sequences in the databases:

>GP:AACT75528 GB:AB000334 orf, hypothetical protein [Escherichia coli]  
 Identities = 143/301 (47%), Positives = 195/301 (64%), Gaps = 21/301 (6%)

45

Query: 1 MKTDDLRESQVEDRRGQSSG-SFGGGLGGLLQLLFSRGWTKLVILLILLVWG-- 57  
 M+ RES VEDRR S G S GG G +L +G L++L+++LV G  
 Sbjct: 1 MRWQRRRESNVDERRSSSGPSMGPGF-----RLPSGKG-----LILLIVLVVAGVY 50

50

Query: 58 GGLGGLGKRPSTNNAYQSSQVTRINGDKASQBVSVKVPASTDYWKTFREKG 117  
 L G+G+ G+P S QG++ N D+A++ F S +A+TSD W + F + G  
 Sbjct: 51 GVNLGLMTGQFVSQQ-----QSTRSISFNDEAAK-----FTSVILATTEDTWGQPFKMG 102

55

Query: 118 IYVHKPTLVLYTATCTACGRGQASSGPFYCPGDQKVYLDISFYNELSTKYGARGDFAMA 177  
 TY +P LV+Y G T+T CG GQ+ GPFYCP D VY+D+SFY+++ K GA GDFA  
 Sbjct: 103 RTYQQPKLVMYRGWTRTGGAGGSGINGPFYCPADGTIVYIDLSFYDMKDKLGADGDFAG 162

60

Query: 178 YVIAHEVGHGHIQNLGIMDNAYASRQGSKAKANLVKLELDADYAGVWYHRYGRNLLGQ 237  
 YVIAHEVGH+Q LGI +Q +A+ N+L+V+ELQD +AG W + +Q GQ+  
 Sbjct: 163 YVIAHEVGHVHQLLGLIEPKVRQLQCNATQAEVNRLSVRMELOADCFAGVWGHSMQQQGL 222

Query: 238 LEKGDIEEAMAAHAGDDTLQETYGKTVDPSTHGTSGKORQRWFRDGYQYQDFEHGDTF 298  
 LE GD+EEA+ AA A+GDD LQ+++ GR VDPSTHGTIS+QR WF RG+ GD +TF  
 Sbjct: 223 LETGLLEALNAAQAIGDRLQQCSQGRVVPDSTHGTISQQRYSWFKRGFDSGDPACQNTF 283

-1545-

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/303 (63%), Positives = 241/303 (79%), Gaps = 5/303 (1%)

```

5   Query: 5   MKIDDLKSDNVEDRRSSGGSGFSSGG-SGLPILQLLLLRGSWKTLVVLIIILLG3GG 63
      MK DDLR+S VEDRR S GSF GG G +LQLL RG WKTKLV+L+LL+GGGG
      Sbjct: 1   MKTDLLRESQVEDRRGQSSGSGFGGGLLQGLLQLLFSRGGWKTLVILLIILLVGGG 60

      Query: 64  LTSIFN---DSSSPSSYSQSVNSVNSATKRGIDFVNKVLGSTRDFWSQSFQTQSGFN 120
      L+ + S++ ++YQS V+R+ + A++BQ+ FV+KV STRD+W++ F+ +G
10   Sbjct: 61  LSGVLGGKPSSTNNAYQSSQVTKNGDKASQKQVSVKVFASDFTWTKYREKGL-T 119

      Query: 121 YKSPKLVLYTNSIQTGCGIGBSASGPFYCSADKKIYLDISFYNELSHKYGATGDFAMAYV 180
      Y +P LVLTY + QT CG G+++SGPFYC D+K+YLDISFYNELS KYGA GDFAMAYV
15   Sbjct: 120 YHKPTLVLYTGATQTACGRGQASSGPFYCPGDQKVIYLDISFYNELSTKYAKGDFAMAYV 179

      Query: 181 IAEHVGHHIQTLEIGIMDKYNNRMHGLTKKANALNVRLQADYTAGVWAHYIRGNILE 240
      IAEHVGHHIQ ELGIMD Y R G +K +AN LNV+LELQADYTAG WA+Y+G+ LLE
      Sbjct: 180 IAEHVGHHIQNELGIMDKYASARQKSKAKANQLNVKLELQADYTAGAWANVYVGGGLE 239

20   Query: 241 QGDPEERMAAAHVGDDTLQETYGKLVFDSFTHGTABQRQSWFNKGFQYGDIGHGDTFS 300
      +GD EEAR AAHAVGDDTLQ+ETYG+ VDSFTHGT++QRQWF++G+QYGD +HGDTFS
      Sbjct: 240 KGDIEERMAAAHVGDDTLQSETYGRVFDSSFTHGTSKQRQRVDRGTQYGDPEHGDITS 299

25   Query: 301 VER 303
      + + +
      Sbjct: 300 IFY 302
  
```

SEQ ID 8800 (GBS404) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 3; MW 62kDa).

30 GBS404-GST was purified as shown in Figure 218, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1401

A DNA sequence (GBSx1486) was identified in *S. agalactiae* <SEQ ID 4301> which encodes the amino acid sequence <SEQ ID 4302>. This protein is predicted to be phenylalanyl-tRNA synthetase beta chain (pheT). Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence
  
```

```

40   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2617 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

45 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CB14823 GB:Z99118 phenylalanyl-tRNA synthetase (beta subunit)
[Bacillus subtilis]
Identities = 376/805 (46%), Positives = 523/805 (64%), Gaps = 6/805 (0%)
  
```

```

50   Query: 1   MLVSYKWLKELVDVD-VTTAEIAEKNSITGIRVEGVSFPAEGLSKLVGVGHVSCROVDIT 59
      M VSYKWL+ VD+ + A LAEK++ GIEVEG+E EG+ +V+GH++ R P+
      Sbjct: 1   MFSYKWLDELVDLKGMDPAVLAEKTIKAGIRVEGIEYKGEIGKGVVIGHVLREGHENA 60

      Query: 60  H-LHLQVDVTGDEDLRQVVGAPNVKTIKNIVIVAFPGARIADNYIKIKGKIRGMSGLMI 118
      L+ C VD G + Q++CGAPNV G V VA GA + N+KIKK K+RG RS GMI
55   Sbjct: 61  DKLNKCLVDIGASAPVQILCGAPNVDRGQKVAATVGAVLGNFIIKIKKIKRGRSNGMI 120
  
```

-1546-

Query: 119 CSLQELGLSESIIPKEFSDGIQILPEGAIPGDISVYLDDDEIIEKLSITTPRADALSNR 178  
CSLQELG+ ++ KE++GI + P A G + L LDD I+EL +TPRADADA+M

5 Sbjet: 121 CSLQELGIESKLVAGYAEIGFVFPNDAAETGSDALAAQLDDALIELGLTPRRADSMQML 180

Query: 179 GVAHEVAALYGVKKVHFEKNLIEAERAADKISVVLIESDKVLS-YSARIVKMTVPASPQ 237  
GVA+EVAAI +V + + +E+A+D ISV IE + Y+A+I+KMTV+APSP

10 Sbjet: 181 GVAHEVAALYDTEKLEQTDYPAASEQASDYSVKIEDQSANELYTAKIKMTVAPSP 240

Query: 238 WLGNKLMNAGIRFINNVVDVINYVLLTYGQPMHAFDFDKPGDTITVARNAMGEKILITLD 297  
W+Q KLMNAGIRF NNVD+TN+VLL YGQ+HAFD+D+F +V R A E ++TLD

15 Sbjet: 241 WMTKLMNAGIRFHNNDVITFNVLLEYGQPLHAFDYDRFSKEVVRKKAENEMIVITLD 300

Query: 298 GEERDLIADDLVIAVNDQFVALAGVMGCGQSTIEGSSSKTVLEAVFNMGTSIRKTSGLN 357  
+ER L AD LVI + A+AGVMGG +E+ +KT++LEA FNG +RK S L

20 Sbjet: 301 DQERKLSADHLVITNGTKAQAVAGVMGGAASEVOEDTKTILLEAYFNQGVKVRKASGLD 360

Query: 358 LRSESSSRFRKGINVTVSEAMDFAAAMLQELAGGVLSQGVTEGVLETPSVEVSTILGY 417  
LRSESS RFKGL+ V A +AA ++ AGG+VL+G V E L S + +

25 Sbjet: 361 LRSESSVRFEKIDIPARVRLAERAQQLHLVAGGEVLACTVEDHULTIEMNHNVSADK 420

Query: 418 VNIRLGTTLTYTIDKEVFEKLGFAGSGSEVKTFLVPRRRDIAZQADLVEIARIYGYE 477  
V++ LG ++ ++ +++LGF + ++ V VP RR DI I+ DL+EE AR+YGY+

30 Sbjet: 421 VSSVLGLTISKELSIYRKLGFVTGSEADLLVTVPSRRGDITIEEDLIEAARLYGTD 480

Query: 478 KLPTTLPEAGATAGELTSMQRLRRRVRTVAEGAGLSEIITYALTPEKAVOFSTQATMIT 537  
+P+TLPE T G LT Q RR+VR EGAGLS+ ITY+LT +KA F+ + + T

35 Sbjet: 481 NIPSTLPETAGTGTGGLTPYQAKRRKVRFLGAGLSQATYSLANEATAPAESKSLNT 540

Query: 538 ELMWPMIVDRSALRQNVYSGMLDTIAYNVARKNSNLAVYEIGKVFQGTGNKEDLPIEVE 597  
L PM+ +RS LR ++V +LD+++YN+AR+ ++A+YE+G VP ++ P E E

40 Sbjet: 541 VLALMSEERSILRHSLVFNLDGVSYNLARKQDSVALYEGSVF--LTKEEDTKVPCETD 598

Query: 598 TTFPMALTGIVNEEDFQTKSKPVDFFYANGIVEALFIKLK-LDVTFAVQGLASMPGRTA 656  
A+TGL ++ +Q + KPVDFF KGIVE L KL ID Q +HPGRTA

45 Sbjet: 599 RVAGAVTGLKRLKQLNGQEKPKVDFFVNGIVEGLDLKLVDSIEFVQSSRKQHPGRTA 658

Query: 657 TILLGKEIGFVGQVHPQTAKYDIPETYVARINLSTIESQNNQALIFEDITKYPSVSRD 716  
ILL+G IGF+QVHP K+ DI ETYV E++L + + L++ I KYPV+RD

50 Sbjet: 659 NILLNGSLGIFGQVHPSLKELDIEKTYVFELDLHALLASTAPLVTIAPKYPSVTRD 718

Query: 717 IALLASVYSHHDIVRAISTGSKRLTAIKLFDVYAGSHIABZYKSHAYSLTFQNPNDL 776  
IAL+ ++V+ + S I+ +G K L + +EDVY G + + EG KS+A+SL + NP L

55 Sbjet: 719 IALVTDKTYTSQGLSEVIREAGGKLKEVTVFDVYSEHMEGKSSVAFSLQYVNEPCTL 778

Query: 777 TDEEVAKYMEKITSKLVKVNARIR 801  
T+EEV K K+ K+L + A+R

60 Sbjet: 779 TDEEVTAHSEKVLKALDITYQAVLR 803

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4303> which encodes the amino acid  
50 sequence <SEQ ID 4304>. Analysis of this protein sequence reveals the following:

Possible site: 27

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.1283 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

60 Identities = 595/801 (74%), Positives = 687/801 (85%)

Query: 1 MLVSYKWLKELVDVDTPTARLAKMSTTGIEVSEVTEPAEGLSKLVGHVSHCEDVPTD 60  
MLVSYKWLKELVDVDTPTARLAKMSTTGIEVSEVTEPAEGLSKLVGHVSHCEDVPTD 60

65 Sbjet: 6 MLVSYKWLKELVDVDTPTARLAKMSTTGIEVSEVTEPAEGLSKLVGHVSHCEDVPTD 65

-1547-

Query: 61 LHLQCVDTGDDSLRQGVGAGPWNKGINVIVAVPGARIADNYKIKKGKIRGMESLGMICS 120  
 LHLQCVDTGDD+ RQ+VCGAPNWK GI VIVAVPGARIADNYKIKKGKIRGMESLGMICS 125

Sbjct: 66 LHLQCVDTGDETFRQIVGAPNWKAGIKVIVAVPGARIADNYKIKKGKIRGMESLGMICS 125

5 Query: 121 LQELGLSESIIKPEPSDGIQILPEGAIPGDSIFSYLDDDEIIEISITPNRADALSNRQV 180  
 LQELGLS+SIIKPEPSDGIQILPE+PGD+IF YLDDDD IIEISITPNRADALSNRQV 185

Sbjct: 126 LQELGLSDSIIKPEPSDGIQILPEEAIVGDAIFKYLDLDDHIIELSITPNRADALSNRQV 185

10 Query: 181 AHEVAAIYGGKVHPEKNLIEEAERADKISVVIESDKVLSYARSIVNVTVAFSPQWLQ 240  
 AHEVAAIYGGK V F +KNL E + ++ I V I SD VL+Y+R+V+NV V FSPQWLQ 245

Sbjct: 186 AHEVAAIYGGKSVFQPNRLQESQKATSAIEVAIASDNVLTASRVNKNVKSPPQWLQ 245

Query: 241 NKLINAGIRPINNVVDVTNYVLLTYGQFMIAFDPKFGTGTIVARNENKGLITLDGEE 300  
 N LMNAGIRPINNVVDVTNYVLL +GQPMFAFD+DKF+ IVAR A GE L+TLGDE+ 305

Sbjct: 246 NLMNAGIRPINNVVDVTNYVLLTYGQPMFAFDYDKFEDHKIVARAARQGSLSVTLDGSK 305

20 Query: 301 RDLIADDLVIAVNDQPVALAGVWGQSTIEGSSKTVVLEAAVFNKTSIRKTSGRNLRS 360  
 RDL +DLVI V D+PVALAGVWGQ+TEI ++S+TVVLRAAVF+G SIKRKSRLNLRS 365

Sbjct: 306 RDLITFEDLVITVADKPVALAGVWGQATEIDANSQTVVLEAAVFDGKSIRKTSGRNLRS 365

Query: 361 ESSSRFEKGINVDTVSEAMDFAAAMLQELAGGQVLSGQVTEGVLPTEPEVSTTLGTGYVT 420  
 ESSSRFEKGINV TV EA+DFAAAMLQELA GQVLSG V G LPTFPEVST+L YVW 425

Sbjct: 366 ESSSRFEKGINVDTVLEALDFAAAMLQELABGQVLSGHVQAGQLPTEPEVSTSLDLYVW 425

25 Query: 421 RLGTSLTYTDIRFVEKLGFAISGSVVKFTVLPVPRRMDIAIQADLVEIARIYGYKLP 480  
 RLGTSLT+ DI+ +F++LGF ++G E FTV VPRRMD++I ADLVEIARIYGY+KLP 485

Sbjct: 426 RLGTSLTADIQIRIFDLQFGLTGDETSFTVAVPRRMDVSIADLVEIARIYGYDKLP 485

30 Query: 481 TTLPEAGTAGELTSMQRRLRRVKTVAEGAGLSEIITYALTTPKAVQSPQTQINITELM 540  
 TTLPEAG TA ELT Q LRR+VR +AEG GL+HII+YALITPEKAV+F+ ++TELM 545

Sbjct: 486 TTLPEAGTAELITYQALRRKVRGLABGLGLTEIISYALTTPKAVEFAVAPSHTELM 545

Query: 541 NPMVTDRSALRQNVVSGMLDTIAYNVARINSLNAVIEIGKVPBQTNPKEDLPTEVETPT 600  
 NPM+V+RSALRQN+VSGMLDT+AYNVARK SNLA+VEIGK+PBQ NPKEDL EV F 605

Sbjct: 546 NPMVSERSALRQNVVSGMLDTIAYNVARKSNLALYIEIGKIFPEBANKEDLPNEVNEFA 605

40 Query: 601 FALTGLVEEKDFCTYKSPVDFYAKGIVEALPIKLKLDVTFVAQGLASMPGRTATILL 660  
 FA+ GLV +KDFCTI+++ VDF+AKG ++ LF L L V +V K LA+MHPGRTA ILL 665

Sbjct: 606 FAICGLVAQKDFCTYQAVDFYHAKGNLDTLPANLNKLVQYVPTYDLANMHPGRTATILL 665

Query: 661 DGKEIGFVQGVHPQATKQVDIPETTYVAEINLSITIESSQNALIFEDITKVPVSERDIAL 720  
 D + IGFVQGVHP TAK Y IPSTYVAR++++ +E+ + F +ITK+F++RD+ALL 725

Sbjct: 666 DEQVIGFVQGVHPGTAKAYSIPETTYVAELMMALEAALPSDQTPARTIKFPANTRDVAL 725

45 Query: 721 LAESVSHHDIVSAIETSQVKRLTAIKLFDVYAGNIAEYGKSMAYSLTPQNFNDLTDDE 780  
 L VSH IV+ALB+QVKRLT+IKLFDVY G I G KSMAYSLTPQNFNDLTDDE 785

Sbjct: 726 LDREVSHQAVTASAGVKRLTSIKLFDVYEGATIQAGKSMAYSLTPQNFNDLTDDE 785

50 Query: 781 VAKYMEKITKSLVEKNVAREI 801  
 VAKYMEKITK+L E+V AE+R 806

Sbjct: 786 VAKYMEKITKALTEQVAGVVR 806

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 55 Example 1402

A DNA sequence (GBSx1487) was identified in *S. agalactiae* <SEQ ID 4305> which encodes the amino acid sequence <SEQ ID 4306>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

60

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0653 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-1548-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

A related GBS nucleic acid sequence <SEQ ID 9769> which encodes amino acid sequence <SEQ ID 9770> was also identified.

- 5 The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB15205 GB:Z99120 transcriptional regulator [Bacillus subtilis]  
Identities = 60/169 (35%), Positives = 100/169 (58%)

10 Query: 17 ITPCKVGLDNPITLQNTAIEITFRQTFSSHNSSEQLQAFNBSYTLFVLKSEITHARSOTY 76  
+ KK +++ LQ ++IETP TF NS E ++A+ ++ L+ E+++ S +  
Sbjct: 3 VMKKCSREDLQTLQQLSIEITFNDTFKEQNSPENMKAYLESAPNTEQLEKELSNMSQFF 62

15 Query: 77 FYILDTDLVGYLKVWNGSQQTEKDLKAFETICRIYLLDAYQCGGIGKATFEFALDLAYS 136  
F+Y D ++ GF+KVN Q+E+ ++ EI+RIY+ ++Q G+GK A+++A +  
Sbjct: 63 FIYFDHEIAGYVKVWIDDAQSEMGASLEIERTIYINSFQKHGLKGLINKATETALER 122

20 Query: 137 GLDNANLGVNPFNHKAQAFYAKYGFKEKFSERQFVGDVDTDWLRKSL 185  
NLGVNE N A AFY K GF + H F +GD+ TD ++ K+L  
Sbjct: 123 NKCHNLGVNFKENALAFYKMGFVQGTGHSFPMGDEQTDLMKARTL 171

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1403

- 25 A DNA sequence (GBSx1488) was identified in *S.agalactiae* <SEQ ID 4307> which encodes the amino acid sequence <SEQ ID 4308>. This protein is predicted to be phenylalanyl-tRNA synthetase (alpha subunit) (pheS). Analysis of this protein sequence reveals the following:

Possible site: 45  
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3937 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 9339> which encodes amino acid sequence <SEQ ID 9340> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB14824 GB:Z99118 phenylalanyl-tRNA synthetase (alpha subunit)  
[Bacillus subtilis]  
Identities = 209/338 (61%), Positives = 270/338 (79%), Gaps = 2/338 (0%)

40 Query: 1 MKISTQELKEM-TGNHTKRLQDLRVQVLSGKGSLSITLKLKLDLNDLRPVVGKQVNEV 59  
+K QE L+++ + K + D+RVQ LSGKG +TE+L+G+ LS + RP +G NEV  
45 Sbjct: 5 LQLEQEALEQVQENASSLKVNDIRVQVLSGKGSPITKVLKNGKLSAERPKMGALANEV 64

50 Query: 60 RDILTKAFEDQAKVVEAAKIQALESESVDVILPGRQMTLGRHRLVLTQTSERIEDIFLGM 119  
R+ + A ++ +E +++ +L +++DVILEG + +G RH LT EEIED+P+GM  
Sbjct: 65 RERTANALADNKELEERERKQKLAGQTDVILPQNPVAVGGRHPTLVIERIEDLFIQM 124

55 Query: 120 GFQVVDGFEVEKDYNNFERMNLPKDHPARDMDQTFYITEILLRTHTSPVQARTNDQHD 179  
G+ V +G EVE DYNNFE +NLPK+HPARDMDQ+FYITEE L+RT TSPVQ RTN++H+  
Sbjct: 125 GYTVREGPEVETDYNNFESNLPEKHPARDMDQSFYITEKLMRTOTSPVQRTMEKEH- 183

Query: 180 SKGPLAMISFGRVFRDDTDATISHQPHQIEGLVVGGENISMGDLKGLTQLQLISGNPFAER 239

-1549-

KGP+K+I PG+V+RD DDATHSHQF QIEGLVV +NISM DLKGTLL+L++KMPG +R  
 Sbjct: 184 GKGPKVLCPCGKVVRRDNDATHSHQFMQLBGLVVDKNISMSDLKGTLLVAKKMPGQDR 243

Query: 240 KIRLRPSYFPPTFESVEVDVSCPKGGKGCNVCKQTGWIRILGAGMVHPSVLEMSGIDSE 299  
 +IRLRPS+FPPTFESVEVDV+CPKGG GC+VCK TGMIRILGAGMVHPS+VL+M+G D +  
 Sbjct: 244 KIRLRPSYFPPTFESVEVDVTCFKGGKGCNVCKQTGWIRILGAGMVHPSVLEMSGIDSE 303

Query: 300 KYSGFAPGLQGERIAMLRYGINDIRGFYQGDVRPTDQF 337  
 +Y GFAPG+G BRIAML+YGI+DIR FY DVRF QF  
 Sbjct: 304 EYQGFAPGVMGVERIAMLRYGINDIRHFTYNDVRFISQF 341

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4309> which encodes the amino acid sequence <SEQ ID 4310>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2806(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 305/337 (90%), Positives = 327/337 (96%)

Query: 1 MKISTQKLEKMENTGNHTKELQDLRVQLGKGSILHELLKGLKDLNDLRFVVGKQVNEVR 60  
 +K T E L+ +TGNTKELQDLRV VLGGKGLTELLKGLKDLNDLRFVVGKQVNEVR  
 Sbjct: 36 LKTKTLETQLSTGNHTKELQDLRVVLGGKGLTELLKGLKDLNDLRFVVGKQVNEVR 95

Query: 61 DLLTKAFEBQAKVBAAKIQAQLSESEVDVTLPRQMTLGHRHVLTQTSSEIEDIPLMG 120  
 D+LTKAFEBQAK+VBAAKIQAQL+ES+DVTLPRQMTLGHRHVLTQTSSEIEDIPLMG  
 Sbjct: 96 DLLTKAFEBQAKVBAAKIQAQLDRESIDVTLPRQMTLGHRHVLTQTSSEIEDIPLMG 155

Query: 121 FOIVDGFVEKDYNYNFERMNLPKDHPARMQDTFYITEILLRTHTSVPQARTDQHDPS 180  
 FO+VDGFVEKDYNYNFERMNLPKDHPARMQDTFYITEILLRTHTSVPQART+DQHDPS  
 Sbjct: 156 FOIVDGFVEKDYNYNFERMNLPKDHPARMQDTFYITEILLRTHTSVPQARTDQHDPS 215

Query: 181 KGFLAMISPGRVFRRTDDATHSHQFHOIEGLVVGNIHSMGDLKGTLLQLISQKMPQAEK 240  
 KGFLAM+SPGRVFRRTDDATHSHQFHOIEGLVVGNIHSMGDLKGTLL+I +KMPG ER  
 Sbjct: 216 KGFLAMISPGRVFRRTDDATHSHQFHOIEGLVVGNIHSMGDLKGTLLMIKKMPQAEK 275

Query: 241 IRLRPSYFPPTFESVEVDVSCPKGGKGCNVCKQTGWIRILGAGMVHPSVLEMSGIDSEK 300  
 IRLRPSYFPPTFESVEVDVSCPKGGKGCNVCK+TGMIRILGAGMVHPSVLEMSG+D+++  
 Sbjct: 276 IRLRPSYFPPTFESVEVDVSCPKGGKGCNVCKQTGWIRILGAGMVHPSVLEMSGIDSEK 335

Query: 301 YSGFAPGLQGERIAMLRYGINDIRGFYQGDVRPTDQF 337  
 YSGFAPGLQGERIAMLRYGINDIRGFYQGD RF++QF  
 Sbjct: 336 YSGFAPGLQGERIAMLRYGINDIRGFYQGDVRFISQF 372

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1404

A DNA sequence (GBSx1489) was identified in *S.agalactiae* <SEQ ID 4311> which encodes the amino acid sequence <SEQ ID 4312>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2834(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1550-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1405

A DNA sequence (GBSx1490) was identified in *S.agalactiae* <SEQ ID 4313> which encodes the amino acid sequence <SEQ ID 4314>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2762(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1406

A DNA sequence (GBSx1491) was identified in *S.agalactiae* <SEQ ID 4315> which encodes the amino acid sequence <SEQ ID 4316>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8801> which encodes amino acid sequence <SEQ ID 8802> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1  Crend: 5
McG: Discrim Score: 10.13
GVH: Signal Score (-7.5): -5.07
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 1 value: -6.79 threshold: 0.0
      INTEGRAL Likelihood = -6.79 Transmembrane 8 - 24 ( 6 - 27)
      PERIPHERAL Likelihood = 6.26 258
      modified ALOM score: 1.86

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```



-1551-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA36134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
Identities = 154/232 (66%), Positives = 180/232 (77%), Gaps = 1/232 (0%)

5   Query: 41  KTVSGTGRRLSSSVLTSTNVKKQLGTINLWNGSGAPIINQNKTLDAKVSAPYALNEIK 100
      K S PS+ L+ESVLT VK Q+ ++ WN SGAPI+N NKT+L+AKVSS PYA N+ K
      Sbjct: 43  KQAEARPSQALAESEVLTDAVKSQIKGLEWNGSGAPIVKNKTNLDAKVSSEKPYADNKT 102

10  Query: 101  KVNNOIVPTKANALLTKATQYRNREETENGRTYMKPAGNEHQINGLKGSYNAVDRGHLL 150
      V + VPT ANALL+KATHQY+NR+ETUNG T W P GWEQ+ LKGSY HAVDRGHLL+
      Sbjct: 103  TVGKETVPTVANALLSKATQYKNRKKTGNGSTWTFPGWEQVKNLKGSTHAVDRGHLL 162

      Query: 161  GYALVGSRLGFDASTSNPKNIATQAAWNAQNSMOSTGQNYETLVKRALDRHKTVRYRV 220
      GYAL+G L GFDASTSNPKNTA Q AWAQA + STGQNYE+ VRKALD++K VRYRV
15  Sbjct: 163  GYALIGSLGFDASTSNPKNIATQTAWNAQNEYSTGQNYEYKVRKALDQNKVRYRV 222

      Query: 221  TLIV-DRDNLSSGSHIEAKSSDGLFNFVTFVQSGLEFDTATGKVKQTK 271
      TL Y ++L+ S S IEAKSSDG LRFNV +FVQ GL DY TG+V T+
20  Sbjct: 223  TLIVASNEDLVPSASQIEAKSSDGLFNFVTFVQSGLEFDTATGVEVTVQ 274
```

There is also homology to SEQ IDs 368 and 1302.

SEQ ID 8802 (GBS285) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 6; MW 32kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 7; MW 57.5kDa).

25 GBS285-GST was purified as shown in Figure 208 (lane 7) and Figure 225 (lane 8).

GBS658 was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 134 (lane 8 & 9; MW 27kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 30 Example 1407

A DNA sequence (GBSx1492) was identified in *S. agalactiae* <SEQ ID 4317> which encodes the amino acid sequence <SEQ ID 4318>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

35  ----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 45 Example 1408

A DNA sequence (GBSx1493) was identified in *S. agalactiae* <SEQ ID 4319> which encodes the amino acid sequence <SEQ ID 4320>. This protein is predicted to be UDP-N-acetylglucosamine (murA). Analysis of this protein sequence reveals the following:

-1552-

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.1814 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9767> which encodes amino acid sequence <SEQ ID 9768>  
 10 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB15693 GB: Z99122 UDP-N-acetylglucosamine  
 1-carboxyvinyltransferase [Bacillus subtilis]  
 Identities = 248/423 (58%), Positives = 323/423 (75%), Gaps = 5/423 (1%)

- 15 Query: 5 MDKIIVBGGQTQIQGVVIRGAKNAVLPILAAATILPSQGRKLLTNVPILSDVFTMNVVR 64  
 M+KIIV GGG +L G V +BGAKNAVLP+AA+L S+ K+++ +VP LSDV+T-N V+R  
 Sbjct: 1 MEKIIVGGG-KIANGTVKVSAGAKNAVLPVIAASLLASSEKSVICDPTLSDVITINEVLR 59
- 20 Query: 65 GLDIQVDFNCKKEILVDASGDILDVAPTEFVSQMRASIVVLGPILARNGHAKVSPGGC 124  
 L V F + E+ V+AS + AP+E+V +MRAS++V-GP+LAR GHA+V++PGGC  
 Sbjct: 60 HLGADVHF--EINNEVTNRSYALCTEAPFEYVRKMRASVLMGPLLARTGARVALPGGC 117
- 25 Query: 125 TIGSRPIDLHLKGLKLEAMGATITTCNGSDITAAQS-KLNGANIYMDFFSVGATQNLMAAATL 183  
 TIGSRPID HLKGL EAMGA I G I A+ +L+GA IY+DFFSVGAT+HL+MAA L  
 Sbjct: 118 AIGSRPIDCHLNGFEAMGAETKVGNGFIEAEVKGRLQGAKIYLOFFSVGATENLMAAAL 177
- 30 Query: 184 ASGTTTIENAAEPEIVDLAQLLNNGAKVKGAGTEITLIIGVDALHSHDHVVQDRIEA 243  
 A GTT+EN A+EPEIVDLA +N MG K++GAGT T+ I GV+ LHG +H ++ DRIEA  
 Sbjct: 178 AGTTTTLENVAKEPEIVDLNANINGMGKIRGAGTGTIKIBGVEKLGSVKHIIIPRIEA 237
- 35 Query: 244 GTFMVAARMTSGNVLVKDAIMEHNRPLISKMEMKVEVSEEDGIRVKADTKKIKVTVK 303  
 GTFMVAAA+T GNVLVK A+ EH LI+K+ EMGV ++E +G+RV E+LKG+ +K  
 Sbjct: 238 GTFMVAALTEGNVLVKGAVPEHLTSLIAKMEGVVTKDGEGLRV-IGPKELKPIDIK 296
- 40 Query: 364 GCRALQAPAMSTDLRASAALILAGNVAQGGTVVQQLTHLDGYYQFHEKLAALGANIKR 423  
 G LQA V +TDLRA AALILAG+VA+G T V +L HLDGYY FH+KLAALGA+T+R  
 Sbjct: 357 GPVQLQAEVAATDLRAGAALILAGLVAGCHTVRLVHLDGYYDFFHQKLAALGADIER 416
- 45 Query: 424 VSE 426  
 V++  
 Sbjct: 417 VND 419

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4321> which encodes the amino acid  
 sequence <SEQ ID 4322>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL likelihood = -3.03 Transmembrane 377 - 393 ( 376 - 394)

----- Final Results -----

- 55 bacterial membrane --- Certainty=0.2211 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: CAB15693 GB: Z99122 UDP-N-acetylglucosamine  
 1-carboxyvinyltransferase [Bacillus subtilis]  
 Identities = 248/423 (58%), Positives = 318/423 (74%), Gaps = 5/423 (1%)

-1553-

Query: 1 VDKIIIEGGQTRLBGEVVIIGAKNAVLPFLAASILPSGKKTILRNVPILSDVPTMNVNR 60  
 ++KII+GGQ+L G V +EGAKNAVLP++AAS+L S+ K+++ +VP LSDV+T+N V+R  
 Sbjct: 1 MEKIIIVRGQ-KLNGTVKVEGAKNAVLPVIAASLASREKSVICDPTLSDVPTVINELR 59

Query: 61 GLDIRVDNFSAANEITVDASGHILDEAPYEVVSQMRASIVVLGPILARNHAKVSMFGGC 120  
 L V F NE+TV+AS + RAP+EV +MRAS++V+GP+LAR GRH+V++PGGC  
 Sbjct: 60 HLGADVHFEN--NEVTVNASYALCTRAPPEYVRKMRASVLMGFLAKTGHARVALRGSC 117

Query: 121 TIGSRPILHLKGLKLEAMGATITQGGDITQAQD-RLOGAMITYMDFPSVGATQNLMAATL 179  
 TIGSRPI+HLKG KANGA I G I A+ RLQGA IY+D+PSVGAT+NI+MAA L  
 Sbjct: 118 AIGSRPIDHLKGFPMGAEIKVGNIGFIAEVKRLQCAKIYLDPPSVGATENLMAAL 177

Query: 180 ADGVTTIENAAAREFEIVDLAQFLNKMGRIRAGAGTEITITGVTHLGRVHDDVQDRIR 239  
 A+G TT+EN A+REFIVDLA ++N MG +IRGAGT T+ I GV L GV+H ++ DRIEA  
 Sbjct: 178 ABGTTTLENVAKFEFEIVDLANYINGMGKIRGAGTGTITIKIGVEKLHGVKIHIIPORIPA 237

Query: 240 GTFMVAAAMTSGNVLRDAVMEHNRPLISKLMEMGVSTEEYGIKRVQANTPKLKPVTVK 299  
 GTFMVAAA+T (NVL++ AV EH LI+K+ EMGV++ +E G+RV +LKP+ +K  
 Sbjct: 238 GTFMVAAAITSGNVLVKGAVPEHLTSLIAKMEEMGVITKDBOGLKV-IGPKELKPIDIK 296

Query: 300 TLPHGPFPTMQAQFTALMAVVGNESTMVETVFNRPCHLEMRMRGLQTEILRDTAMIH 359  
 T+PHGPFPTMQ+Q AL+ +G S + ETVPENRF H EE RM +I + +I+  
 Sbjct: 297 TMHPGPFPTMQSGMALLRASGTSMTTETVFNRPCHLEMRMRGLQTEILRDTAMIH 356

Query: 360 GGRQLQAPVMSTDLRASAALILITGIVAQGVTVNNVLHLDGRGYQFHEKLAKLGATIR 419  
 G QLQGA V +TLRRA AALL G+VA+G T V L HLDGRGY FH+KLA LGA I R  
 Sbjct: 357 GPVQLQGAZVAATDLRAGAALILAGLVAKGHTVTELKHLDRGVYDFHQLAALGADIKR 416

Query: 420 SSE 422  
 ++  
 Sbjct: 417 VND 419

An alignment of the GAS and GBS proteins is shown below.

Identities = 363/422 (86%), Positives = 391/422 (92%)

Query: 5 MDKIIIEGGQTRLGQGVVIGAKNAVLPFLAASILPSGKKTILRNVPILSDVPTMNVNR 64  
 +DKII+EGGQ+L+G+VVIIGAKNAVLPFLA++ILPS+GKT+L NVPI LSDVPTMNVNR  
 Sbjct: 1 VDKIIIEGGQTRLBGEVVIIGAKNAVLPFLAASILPSGKKTILRNVPILSDVPTMNVNR 60

Query: 65 GLDIQVDFNCKEILVDASGHILDVAPYEVVSQMRASIVVLGPILARNHAKVSMFGGC 124  
 GLDI+VDPN EI VDRSG ILD APYB+VSQMRASIVVLGPILARNHAKVSMFGGC  
 Sbjct: 61 GLDIRVDNFSAANEITVDASGHILDEAPYEVVSQMRASIVVLGPILARNHAKVSMFGGC 120

Query: 125 TIGSRPIDHLKGLKLEAMGATITQGGDITQAQKILKGANIYMDFPSVGATQNLMAATLA 184  
 TIGSRPI+HLKGLKLEAMGATITQ GGDITQAQA++L+GA IYMDFPSVGATQNLMAATLA  
 Sbjct: 121 TIGSRPIDHLKGLKLEAMGATITQGGDITQAQDRIGAMITYMDFPSVGATQNLMAATLA 180

Query: 185 SGTTIENAAAREFEIVDLAQFLNKMGRIRAGAGTEITITIGVDALHGRVHDDVQDRIR 244  
 G TTIENAAAREFEIVDLAQ INKGA++GAGTETIT GV L G EHDVQDRIRAG  
 Sbjct: 181 DGVTIENAAAREFEIVDLAQFLNKMGRIRAGAGTEITITIGVTHLGRVHDDVQDRIRAG 240

Query: 245 TFMVAAAMTSGNVLRDAVMEHNRPLISKLMEMGVSTEEYGIKRVQANTPKLKPVTVK 304  
 TFMVAAAMTSGNVL++DA+MEHNRPLISKLMEMGV V+EE GIRV+A+T KLEKPVTVK  
 Sbjct: 241 TFMVAAAMTSGNVLRDAVMEHNRPLISKLMEMGVSTEEYGIKRVQANTPKLKPVTVK 300

Query: 305 LPHGPFPTMQAQFTALMAVVGNESTMIETVPENRFCHLEMRMRGLQTEILRDTAMIH 364  
 LPHGPFPTMQAQFTALMAVVGNESTMIETVPENRFCHLEMRMRGLQ+EILR+TAMIH  
 Sbjct: 301 LPHGPFPTMQAQFTALMAVVGNESTMVETVPENRFCHLEMRMRGLQSKILRSTAMIH 360

Query: 365 GRALQGAQVMSTDLRASAALILRGMVAQGVTVGOLTHLDGRGYQFHEKLAKLGATIR 426  
 GR LQGAQVMSTDLRASAALIL G+VAQ T+V L HLDGRGYQFHEKLAKLGATIR  
 Sbjct: 361 GRQLQGAQVMSTDLRASAALILITGIVAQGVTVNNVLHLDGRGYQFHEKLAKLGATIR 422

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1554-

**Example 1409**

A DNA sequence (GBSx1494) was identified in *S. agalactiae* <SEQ ID 4323> which encodes the amino acid sequence <SEQ ID 4324>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2096 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA23756 GB:AB009314 proton-translocating ATPase, epsilon
subunit [Streptococcus bovis]
Identities = 102/138 (73%), Positives = 121/138 (86%), Gaps = 1/138 (0%)

Query: 1 MAQLTVQVVTFDGIKYDHHASLITVTPDGMGILPGHINLIAPLVHQMKNRSHQEG- 59
M Q+TVQVVTFDGIKYDHHASLITVTPDGMGILPGHINLIAPLVHQMKNRSHQEG- 59
Sbjct: 1 MTFMTVQVVTFDGIKYDHHANFISVKTTPDGMGILPGHINLIAPLVHQMKNRSHQEG- 60

Query: 60 VDWAVNGGIIIEVNEQVTVIVADSAERARDIDLNRRAERAKERAERALEKAQTTONIDEMR 119
VDWAVNGGIIIE++ VTIIVADSAER RDID++RAERAK RAER IE+AQ+T +IDB+R
Sbjct: 61 VDWAVNGGIIIEIKDLVTVIVADSAERERDIDVSRRAERAKIRAERKLEQAQSTHIDISVR 120

Query: 120 RAEVALRRAINRISVGK 137
RA+VALRRA+NRISVG K
Sbjct: 121 RAEVALRRAINRISVGK 138

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4325> which encodes the amino acid sequence <SEQ ID 4326>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2539 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 100/138 (72%), Positives = 119/138 (85%), Gaps = 1/138 (0%)

Query: 1 MAQLTVQVVTFDGIKYDHHASLITVTPDGMGILPGHINLIAPLVHQMKNRSHQEG- 59
M Q+TVQVVTFDGI+YDHHASLITVTPDGMGILPGHINLIAPLVHQMKNRSHQEG- 59
Sbjct: 1 MTFMTVQVVTFDGIKYDHHAKFISVKTTPDGMGILPGHINLIAPLVHQMKNRSHQEG- 60

Query: 60 VDWAVNGGIIIEVNEQVTVIVADSAERARDIDLNRRAERAKERAERALEKAQTTONIDEMR 119
VDWAVNGGIIIE++ VTIIVADSAER RDID++RAERAK RAER +A+TT NIDE+R
Sbjct: 61 VDWAVNGGIIIEIKDLVTVIVADSAERERDIDVSRRAERAKIRAERKLEQAQSTHIDISVR 120

Query: 120 RAEVALRRAINRISVGK 137
RA+VALRRA+NRISVG K
Sbjct: 121 RAEVALRRAINRISVGK 138

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1555-

**Example 1410**

A DNA sequence (GBSx1495) was identified in *S.agalactiae* <SEQ ID 4327> which encodes the amino acid sequence <SEQ ID 4328>. Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein is similar to the beta subunit of the S.mutans ATPase:

>GP:AD13383 GB:U31170 ATPase, beta subunit [Streptococcus mutans]
Identities = 435/466 (93%), Positives = 455/466 (97%)

Query: 1  MSSGKIAQVVGPVVDVVFASGDKLPENNNALIVYKNGDKSQKVLEVALELGDGLVRTIA 60
MS+GKIAQVVGPVVDV FA+ DKLPEIDNAL+VYK+GDKSQ++VLEVALELGDGLVRTIA
Sbjct: 1  MSTGKIAQVVGPVVDVAFATDDKLPEINNALIVYKDGDKSQRIVLEVALELGDGLVRTIA 60

Query: 61  MESTDGLTRGLEVDITGRAISVPVKGDTLGRVFNVLGDCAIDLEEPFADAERQPIHKKAP 120
MESTDGLTRGLEV DTGRAISVPVKG+TLGRVFNVLGD IDL++PFADAERQPIHKKAP
Sbjct: 61  MESTDGLTRGLEVDTGRAISVPVKGDTLGRVFNVLGDTIDLDKPFADAERQPIHKKAP 120

Query: 121  SFDELSTSSIELETGKIVKIDLLAPYLEGGKVLFGGAGVGKTVLIQELTHNIAQEHGGIS 180
SFD+LST+SELETGKIVKIDLLAPYLEGGKVLFGGAGVGKTVLIQELTHNIAQEHGGIS
Sbjct: 121  SFDELSTSSIELETGKIVKIDLLAPYLEGGKVLFGGAGVGKTVLIQELTHNIAQEHGGIS 180

Query: 181  VFTGVGERTREGNDLYWEMKESGVTEKTMVFGQMNPEPPGARMEVALTGLTIAEYFRDVE 240
VFTGVGERTREGNDLYWEMKESGVTEKTMVFGQMNPEPPGARMEVALTGLTIAEYFRDVE
Sbjct: 181  VFTGVGERTREGNDLYWEMKESGVTEKTMVFGQMNPEPPGARMEVALTGLTIAEYFRDVE 240

Query: 241  GQDVLLFIDNIFRPTQAGSEVSALLGRMPSAVGYQPTLATMGQLQERITSTKKGVSVTSI 300
GQDVLLFIDNIFRPTQAGSEVSALLGRMPSAVGYQPTLATMGQLQERITSTKKGVSVTSI
Sbjct: 241  GQDVLLFIDNIFRPTQAGSEVSALLGRMPSAVGYQPTLATMGQLQERITSTKKGVSVTSI 300

Query: 301  QAIYVPADDDYTPAPATAFAHLOSTINLERKLTQMGIYPAVDPLASSRAL+PEIVG E H 360
QAIYVPADDDYTPAPATAFAHLOSTINLER+LTQMGIYPAVDPLASSRAL+PEIVG EH
Sbjct: 301  QAIYVPADDDYTPAPATAFAHLOSTINLERRLTQMGIYPAVDPLASSRAL+PEIVGQEH 360

Query: 361  YVATEWQVRVLQRYRELQDIIAILGMDLSDEKTLVGRARRIQFFLSQNFNVAESTFTQ 420
Y+VATEWQ VLQRYRELQDIIAILGMDLSDEKTLVGRARRIQFFLSQNFNVAESTFTQ
Sbjct: 361  YDVATEWQVRLQRYRELQDIIAILGMDLSDEKTLVGRARRIQFFLSQNFNVAESTFTQ 420

Query: 421  FGSYVPVEETVRGPKKILDGKHQIPEDAFRMWGQIEDVITAKEM 466
FGSYVPV ETVRGPKKIL+GK+D++PEDAFR VG IEDV+ KAKKM
Sbjct: 421  FGSYVPVAETVRGPKKILGKYDGLAFDAFNSVGAIEDVVEKAKKM 466

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4329> which encodes the amino acid sequence <SEQ ID 4330>. Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0275 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 440/468 (94%), Positives = 456/468 (97%)

Query: 1  MSSGKIAQVVGPVVDVVFASGDKLPENNNALIVYKNGDKSQKVLEVALELGDGLVRTIA 60

```

-1556-

- MS8GKIAQVVGVDV+ FASGDKLFEINNALIVK+ DK QK+VLEVALELGG+VETIA  
 Sbjct: 1 MS8GKIAQVVGVDVFMFASGDKLFEINNALIVKDSKQKIVLEVALELGGNVRTIA 60
- 5 Query: 61 MESTDGLTRGLEVLDTGRAISVPVGKDTLGRVFNVLGDAIDLEEFPAEDARQPIHKKAP 120  
 MESTDGLTRGLEVLDTGRAISVPVGK+TLGRVFNVLG+ IDLEEFPAED+RQPIHKKAP  
 Sbjct: 61 MESTDGLTRGLEVLDTGRAISVPVGKDTLGRVFNVLGGETIDLEEFPAEDVRDQPIHKKAP 120
- 10 Query: 121 SPDELSTSSSEILGTGIKVIDLLAPYLKGGKVGGLFGGAGVGKTVLIQELIHNLIAQEHGGIS 180  
 SPDELSTSSSEILGTGIKVIDLLAPYLKGGKVGGLFGGAGVGKTVLIQELIHNLIAQEHGGIS  
 Sbjct: 121 SPDELSTSSSEILGTGIKVIDLLAPYLKGGKVGGLFGGAGVGKTVLIQELIHNLIAQEHGGIS 180
- 15 Query: 181 VFTGVGERTRGNDLYWEMKESGVIEKTAMVFSQNEPFGARMRVALTGLTIAEYFRDVE 240  
 VFTGVGERTRGNDLYWEMKESGVIEKTAMVFSQNEPFGARMRVALTGLTIAEYFRDVE  
 Sbjct: 181 VFTGVGERTRGNDLYWEMKESGVIEKTAMVFSQNEPFGARMRVALTGLTIAEYFRDVE 240
- 20 Query: 241 GQDVLLFDINLPRFTQMGSEVALLGRMPASVGYQPTLATMGQLQERITSTKKGSVTSI 300  
 GQDVLLFDINLPRFTQMGSEVALLGRMPASVGYQPTLATMGQLQERITST+KKGSVTSI  
 Sbjct: 241 GQDVLLFDINLPRFTQMGSEVALLGRMPASVGYQPTLATMGQLQERITSTKKGSVTSI 300
- 25 Query: 301 QAIYVPADYTDPAAPATAFAHLDSITNLERKLTQMGIVPAVDPLASSRALTPAIWGDEN 360  
 QAIYVPADYTDPAAPATAFAHLDSITNLERKLTQMGIVPAVDPLASSRAL+PEIVG+EN  
 Sbjct: 301 QAIYVPADYTDPAAPATAFAHLDSITNLERKLTQMGIVPAVDPLASSRALSPETVGEEN 360
- 30 Query: 361 YVATEVQRVLQRYRELQDIIALLGWDELSDDEKTLVGRARRIQFFLSQNFVAETFTGQ 420  
 Y VATEVQRVLQRYRELQDIIALLGWDELSDDEKTLVGRARRIQFFLSQNFVAE FTG  
 Sbjct: 361 YAVATEVQRVLQRYRELQDIIALLGWDELSDDEKTLVGRARRIQFFLSQNFVAEQFTGL 420
- Query: 421 PGSYVPVEETVIRGPKELIDGRHDQIPSDAFRNVGGIEDVIAKAEKMY 468  
 PGSYVPV +TVIRGPKEL+QK+D+PSDAFR VG IEDVI KAEK +  
 30 Sbjct: 421 PGSYVPADTVIRGPKELLEGYDELSDAFRSVGIEDVIAKAEKMGF 468
- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.
- Example 1411**
- 35 A DNA sequence (GBSx1496) was identified in *S. agalactiae* <SEQ ID 4331> which encodes the amino acid sequence <SEQ ID 4332>. Analysis of this protein sequence reveals the following:
- Possible site: 31  
 >>> Seems to have no N-terminal signal sequence
- 40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1889 (Affixmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- 45 The protein has homology with the following sequences in the GENPEPT database.
- >GP:BA23754 GB:AB009314 proton-translocating ATPase, gamma subunit  
 [Streptococcus bovis]  
 Identities = 252/293 (86%), Positives = 278/293 (94%), Gaps = 2/293 (0%)
- 50 Query: 1 MAGSLSEIKDKILSTERYKITSANQMVSAKLKSEQAARDFQVYASKIRQITITNLKS 60  
 MAGSLSEIK KI+ST+KYS IT ANQMV+AKL KSEQAA+DFQVYASKIRQITT+LKS  
 Sbjct: 1 MAGSLSEIKGKILSTQKSHITGANQMVSAKLKSEQAARDFQVYASKIRQITIDLKS 60
- 55 Query: 61 DLVSGSDNEMLSRPVKNKGYIVTSDKGLVGGYNSKILKAMMDITDYHTENDVYAIIS 120  
 +LV+GS NPM+L+RPVKNKGYIVTSDKGLVGGYNSKILKAMMD I +YH ++ +YAI+  
 Sbjct: 61 ELVSGSKNPMLAARPVKNKGYIVTSDKGLVGGYNSKILKAMMDLIERVH -QGNVYAIIA 119
- Query: 121 IGSVGSDFPKARQMNVSFELAGLEDQSPFDQNGKIIAGVEMYKNELFDLEYVCYNHVN 180  
 IG +G+DFPKARQMNVS FELAGLEDQSPF+QNG IIA+VEMYKNELFDLEYVCYNHVN  
 60 Sbjct: 120 IGGIGADFPKARQMNVSFELAGLEDQSPFDQNGNI IAKSVEMYKNELFDLEYVCYNHVN 179
- Query: 181 SLTSQVRMCQMLPIKELDAREASERVITTFPELSPMRVRVILEQLLPYTESLTYQAIDA 240

-1557-

SLTSQVR+QQLPI ELDA+EA+E+ V +GFELEPNRE+ILQLLPQYTESLIYGA+DA  
 Sbjct: 180 SLTSQVRVQQLPIELDADEAAEBGV-SGFELEPNRE+ILQLLPQYTESLIYGA+VDA 238

Query: 241 KTAHHAAGMTAMQTATDNAKINVINDLTIQYNRQAATQETITEIVAGANALE 293  
 KTAHHAAGMTAMQTATDNAKINVINDLTIQYNRQAATQETITEIVAGANALE  
 Sbjct: 239 KTAHHAAGMTAMQTATDNAKINVINDLTIQYNRQAATQETITEIVAGANALE 291

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4333> which encodes the amino acid sequence <SEQ ID 4334>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1969 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 251/293 (85%), Positives = 275/293 (93%), Gaps = 2/293 (0%)

Query: 1 MAGSLSEIKDKILSTEKTSKITSAMQMVSSAKLVKSEQARDQVYASKIRQITTNLLKS 60  
 MAGSLSEIK KI+STEKTSKITSAM+MVSSAKLVKSEQARDQV+YASKIRQITT+LLKS  
 Sbjct: 1 MAGSLSEIKAKIISTEKTSKITSAMRMVSSAKLVKSEQARDQVYASKIRQITTDLLKS 60

Query: 61 DLVSGSDNPMLSRPVKKTGYIVITSDKGLVGGYNSKILK+MD IT+YH + DY IIS 120  
 +L GSDNPM LSRPVKKTGYIVITSDKGLVGGYNSKILK+MD IT+YH + DY IIS  
 Sbjct: 61 ELTIGSDNPMLSRPFVKKTGYIVITSDKGLVGGYNSKILKSVMDMITEYHADG-DYIIS 119

Query: 121 IGSVGSDFPKARGMNVFELRGLDQPSFQVQKIIAQAVEMYNKELFDELYVCYNHVN 180  
 IGSVGSDFPKARGMNVFELRGL DQPSF+QV +II+Q+V+M+ NE+FDLYVCYNHVN  
 Sbjct: 120 IGSVGSDFPKARGMNVAFELRGLDQPSFEQRIIISQVDMFVNEIFDELYVCYNHVN 179

Query: 181 SLTSQVR+QQLPIELDAEASESDRVITGFELEPNRE+ILQLLPQYTESLIYGA+IDA 240  
 SLTSQVR+QQLPI +L A+EA+E+ V TGFLEPNR IL+QLLPQ+TESLIYGA+IDA  
 Sbjct: 180 SLTSQVRVQQLPISDLVADEAAEBGV-TGFLEPNRHIDQLLPQFTESLIYGA+IDA 238

Query: 241 KTAHHAAGMTAMQTATDNAKINVINDLTIQYNRQAATQETITEIVAGANALE 293  
 KTAHHAAGMTAMQTATDNAKINVINDLTIQYNRQAATQETITEIVAGANALE  
 Sbjct: 239 KTAHHAAGMTAMQTATDNAKINVINDLTIQYNRQAATQETITEIVAGANALE 291

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1412

A DNA sequence (GBSx1497) was identified in *S.agalactiae* <SEQ ID 4335> which encodes the amino acid sequence <SEQ ID 4336>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1963 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1558-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1413**

A DNA sequence (GBSx1498) was identified in *S. agalactiae* <SEQ ID 4337> which encodes the amino acid sequence <SEQ ID 4338>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3146(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein is similar to the alpha subunit of the proton-translocating ATPase from *S. bovis*:

```

>GP:BAA23753 GB:AB009314 proton-translocating ATPase, alpha subunit
[Streptococcus bovis] Length = 501
Identities = 482/501 (96%), Positives = 497/501 (99%)

Query: 1 MAINAQEISALIKKIBDFQPNFDVTETGIVTYIGDGIARAGLNNAMSGELLEFSNGAY 60
      20 MAINAQEISALIKKIB+QPNFDVTETG+VTYIGDGIARAGLNNAMSGELLEFSNGAY+
      Sbjct: 1 MAINAQEISALIKKIBDFQPNFDVTETGVVTTYIGDGIARAGLNNAMSGELLEFSNGAF 60

Query: 61 GMAQNLESNDVGIIILGDFSSIREGDEVVKTGKIMEVFPVGEALIGRVNPLQGVDSGLGE 120
      25 GMAQNLESNDVGIIILGDFSS IREGD VKTGKIMEVFPVGEA+IGRVNPLQGVDSGLG+
      Sbjct: 61 GMAQNLESNDVGIIILGDFSTIREGDEVVKTGKIMEVFPVGEALIGRVNPLQGVDSGLGD 120

Query: 121 IETATRFVETPAPGVMMQRKSVTEPLQTGLKADALVPIGRGRELIIIDRQTGKTSVAI 180
      30 I+TATRFVETPAPGVMMQRKSV EPLQTGLKADALVPIGRGRELIIIDRQTGKTSVAI
      Sbjct: 121 IKTATRFVETPAPGVMMQRKSVSEPLQTGLKADALVPIGRGRELIIIDRQTGKTSVAI 180

Query: 181 DAILNQKQGDWICITYVAIQKQESTVTRTQVETLRKYGALDYTIIVVTASQSPSELLPIAPY 240
      35 DAILNQKQGDWICITYVAIQKQESTVTRTQVETLRKYGALDYTIIVVTASQSPSELL+IAPY
      Sbjct: 181 DAILNQKQGDWICITYVAIQKQESTVTRTQVETLRKYGALDYTIIVVTASQSPSELLYIAPY 240

Query: 241 AGVMAAEFPMYNGKHLIVYDDELKQAVAYRELSLLRRPDPGRAYPGDVFYLHSLRLER 300
      40 AGVMAAEFPMYNGKHLIVYDDELKQAVAYRELSLLRRPDPGRAYPGDVFYLHSLRLER
      Sbjct: 241 AGVMAAEFPMYNGKHLIVYDDELKQAVAYRELSLLRRPDPGRAYPGDVFYLHSLRLER 300

Query: 301 SARVSDALGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLNSGIRPAIDAG 360
      45 SARVSDALGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLNSGIRPAIDAG
      Sbjct: 301 SARVSDALGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLNSGIRPAIDAG 360

Query: 361 SSVSRVGGAAQIQMKRVAGTFLRLDLASVYRELFAPTQFSDILDAATQAKLNRRGRTVEVL 420

```



-1559-

bacterial cytoplasm --- Certainty=0.3654 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

# 5 An alignment of the GAS and GBS proteins is shown below.

Identities = 477/501 (95%), Positives = 490/501 (97%)

Query: 1 MAINAQEISALIKKQIEDFPQNFVDVETGIVTYIGDGIARAGLDNAMSGELLEFSNGAY 60  
 +AINAQEISALIKKQIE+PQNFVDVETGIVTYIGDGIARAGLDNAMSGELLEFS NGAY  
 10 Sbjct: 1 LAINAQEISALIKKQIEFPQNFVDVETGIVTYIGDGIARAGLDNAMSGELLEFSNGAY 60

Query: 61 GMAQNLESNDVGIIILGDFSEIREBGDVVKRTGKIMEVVPGEAMIGRVVNPLGQPDVGLGE 120  
 GMAQNLESNDVGIIILGDFSI REBGDVVKRTGKIMEVVPGEAMIGRVVNPLGQPDVGLG+  
 Sbjct: 61 GMAQNLESNDVGIIILGDFSAIREBGDVVKRTGKIMEVVPGEALIGRVVNPLGQPDVGLGE 120

15 Query: 121 IETTATRPVETPAPGVMMQRKSVPEPLQTGLKALDALVPIGRGQRELIIGDRGTGKTSVAI 180  
 IETT RPVETPAPGVMMQRKSV EPLQTGLKALDALVPIGRGQRELIIGDRGTGKTSVAI  
 Sbjct: 121 IETTGRFPVETPAPGVMMQRKSVSEPLQTGLKALDALVPIGRGQRELIIGDRGTGKTSVAI 180

20 Query: 181 DAILNKGQDMI CIYVAIGQKESTVRTQVETLRKYGALDYTI VVTASASQSPFLFIAPY 240  
 DAILNKGQDMI CIYVAIGQKESTVRTQVETLR+YGALDYTI VVTASASQSPFLFIAPY  
 Sbjct: 181 DAILNKGQDMI CIYVAIGQKESTVRTQVETLRKYGALDYTI VVTASASQSPFLFIAPY 240

25 Query: 241 AGVMAAEFFMYGKHVLIVYDDLRSQAVAYRELSLLRRPFGREAYPGDVLYLHRSLLER 300  
 AGVMAAEFFMY GKHVLIVYDDLRSQAVAYRELSLLRRPFGREAYPGDVLYLHRSLLER  
 Sbjct: 241 AGVMAAEFFMYGKHVLIVYDDLRSQAVAYRELSLLRRPFGREAYPGDVLYLHRSLLER 300

30 Query: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLFNSGIRPAIDAG 360  
 SAKVSD LGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLFNSGIRPAIDAG  
 Sbjct: 301 SAKVSDDLGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLFNSGIRPAIDAG 360

35 Query: 361 SSVSRVGGAAQIKAMKRVAGTLRLDLASYRELEAFTQFSGLDAATQAKLNRRGRTVEVL 420  
 SSVSRVGG+AQIKAMK+VAGTLRLDLASYRELEAFTQFSGLDAATQAKLNRRGRTVE+L  
 Sbjct: 361 SSVSRVGGAAQIKAMKRVAGTLRLDLASYRELEAFTQFSGLDAATQAKLNRRGRTVEIL 420

40 Query: 421 KQPLHKPLPVEKQVILYALTHGFLDDVPVDILAFEEALYDYDFDAHYNDLPETIRTTKD 480  
 KQPLHKPLPVEKQVILYALTHGFLDDVPV+DILAFEEALYDYDF HY++LPETIRTTKD  
 Sbjct: 421 KQPLHKPLPVEKQVILYALTHGFLDDVPVDILAFEEALYDYDFVHYNDLPETIRTTKD 480

45 Query: 481 LPFEAEALCAAIAQKQSQK 501  
 LPFEA LCAAIA+AFK+ S FK  
 Sbjct: 481 LPFEAALCAAIAKPKHSNFK 501

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1414

A DNA sequence (GBSx1499) was identified in *S.agalactiae* <SEQ ID 4341> which encodes the amino acid sequence <SEQ ID 4342>. Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

100 bacterial cytoplasm --- Certainty=0.1896 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 150 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA23752 GB:AB009314 proton-translocating ATPase, delta subunit  
 [Streptococcus bovis]  
 60 Identities = 98/178 (55%), Positives = 127/178 (71%)

-1560-

Query: 1 MNKKTQALIBQYSKSLVEVAIEHKIVEKIQOEVAALIDIFETSELGVLSSLAVSHDEKQ 60  
 M+KKTQAL+BOY+KSLVB+AIE + ++Q E AL+ +FE + L LSSL VS DEK  
 Sbjct: 1 MNKKTQALIBQYAKSLVEVAIEKDKSLAKLQSETALLSVFETMLADLSSLVSRDEKV 60

5 Query: 61 HFVKTLQTSCTSYLVNLFLEVIVQNEREALLYPILKSVQDLIKVNGQYPIQTITVAVALSP 120  
 V+ LQ S S Y+ NPLEVI+QNEREA L IL+ V ++ + Q+ I+TTAVAL+  
 Sbjct: 61 KLVRLQSSSVVMNPLFVILVQNEREAFLLKAILBQVQKDFVIATNQHDIVVTITVAVAL 120

10 Query: 121 EQKERLFDIATKIALPNGQLVEHIDPSIVGGFVFNANKVIDASVRNQLHQPFMKLK 178  
 EQKER+ + K + G+LVE+ID SI+GGFV+N NNVVID S+R QL +FNM LK  
 Sbjct: 121 EQKERILALVAREKPGVKGKLVENIDESILGGFVFNANKVIDTSIRROQLQPFNMMLK 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4343> which encodes the amino acid sequence <SEQ ID 4344>. Analysis of this protein sequence reveals the following:

15 Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1668 (Affirmative) < succ>  
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/178 (48%), Positives = 125/178 (69%)

25 Query: 1 MNKKTQALIBQYSKSLVEVAIEHKIVEKIQOEVAALIDIFETSELGVLSSLAVSHDEKQ 60  
 M KK QALIEQY+KSLVEVA EH ++ +Q +V A+++ F T+ L+ LSS AV H EK  
 Sbjct: 1 MTKKEQALIBQYAKSLVEVAEHHSLDALQADVLAILETFVVTNLQQLSSQAVPHAEKI 60

30 Query: 61 HFVKTLQTSCTSYLVNLFLEVIVQNEREALLYPILKSVQDLIKVNGQYPIQTITVAVALSP 120  
 + L+ + S Y+ NPL +I+QNEREA LY +L+V E+ V+ QY + +T+++ L+  
 Sbjct: 61 KLLTLKNNSVVMNPLFVILVQNEREAYLYQLQAVLNIAIVSNQYDVVTSSLPTE 120

35 Query: 121 EQKERLFDIATKIALPNGQLVEHIDPSIVGGFVFNANKVIDASVRNQLHQPFMKLK 178  
 EQK R+ + K A+ G+L+E +DPS++GGF+++ NNVVID S+R QL FNM LK  
 Sbjct: 121 EQKSRVRAVAKKPAVTAAGRLIEKVDPSLGGFPIISVNNKVIDTSIRROQLQPFNMMLK 178

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 1415

A DNA sequence (GBSx1500) was identified in *S.agalactiae* <SEQ ID 4345> which encodes the amino acid sequence <SEQ ID 4346>. This protein is predicted to be ATP synthase b chain (atpF). Analysis of this protein sequence reveals the following:

Possible site: 33  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]  
 Identities = 103/165 (62%), Positives = 130/165 (78%)

55 Query: 1 MSILNSTTIGDIIIVSGSVLLLFILIKTFANKQITGIPAREQKIANDIDTAQARQQA 60  
 MS LIN T+G+++IV+GS +LL +L+K FAW Q+ IP+ RE+KIA DID AE +RO A  
 Sbjct: 1 MSTLINTSLGKLLIVTGSFIILLLLLVKKEFWQSLIAIFKTRREKIAKIDIDDAENRQNA 60

-1561-

Query: 61 EAFATKREBELSNAKTEANQIIDNAKETGLAKGDQIISRAKTEADRLKEKAHQDIAQNKGA 120  
 + KR+ EL+ AK EA QIIDNAKETG A+ +II+EA EA RLK+KA+QDIA +KA  
 Sbjct: 61 QVLENKRQVLEINQAKDEAAQIIDNAKETGKAQESKIIITRAHEAGRLKDKANQDIATSKA 120

Query: 121 EALADVKGADVADLVLLAEKIMVSNLDKRAQSNLIDSYIKKIGDA 165  
 EAL+ VK +VADL+VLLAEKIM NLDK AQ +LIDSY+ KIGDA  
 Sbjct: 121 EALSSVKADVADLVSVLLAEKIMAKNLDKTAQGDILIDSYLDKIGDA 165

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4347> which encodes the amino acid sequence <SEQ ID 4348>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]  
 Identities = 88/159 (55%), Positives = 122/159 (76%)

Query: 6 GELVGNFILVTGSGVIVLLLLIKKFAWGAIESILQTKSQISROIDQAEQSRLSAQLEAK 65  
 +GN ++VTGS I+LLLL+KKFAW + +I +TR ++I+DID AE SR +AQ LE K  
 Sbjct: 7 GTSIGNLILVTGSPILLLLIKKFAWSQLAIFKTREKIKARDIDDAENSRQNAQVLENNK 66

Query: 66 SQANLDASRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAV 125  
 Q L+ ++ +A++II +AKE G+ Q K++ EA +EA RLK+KA DI SK++A+S+V  
 Sbjct: 67 RQVLEINQAKDEAAQIIDNAKETGKAQESKIIITRAHEAGRLKDKANQDIATSKAELSSV 126

Query: 126 KTEMSDLTVLLAEKIMGANLKDKAQSQSLIDSYLDLGEA 164  
 K +++DL+VLLAEKIM NLDKTAQ LIDSYLD LG+A  
 Sbjct: 127 KADVADLVSVLLAEKIMAKNLDKTAQGDILIDSYLDKIGDA 165

An alignment of the GAS and GBS proteins is shown below.

Identities = 81/156 (51%), Positives = 115/156 (72%)

Query: 10 IGDIIIVGSGVLLLLFILIKTFANKQITGIFEAREBKIANIDITBARQARQAEAFATKRE 69  
 +G+ I+V+GSGV++L +LIK FAW I I + R+Q+I+ DID AEQ+R A+ K +  
 Sbjct: 9 VGNFILVTGSGVIVLLLLIKKFAWGAIESILQTSRQISROIDQAEQSRLSAQLEAKSQA 68

Query: 70 ELSNAKTEANQIIDNAKETGLAKGDQIISRAKTEADRLKEKAHQDIAQNKARALADVKE 129  
 L ++ +A++II +AKE G +GD+++EA EA RLKEKA DI Q+K++A++ VK E  
 Sbjct: 69 NLDASRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAVKTE 128

Query: 130 VADLVLLAEKIMVSNLDKRAQSNLIDSYIKKIGDA 165  
 ++DLTVLLAEKIM +NLDK AOS LIDSY+ LG+A  
 Sbjct: 129 MSDLTVLLAEKIMGANLKDKAQSQSLIDSYLDLGEA 164

SEQ ID 4346 (GBS169) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 6; MW 18kDa).

The GBS169-His fusion product was purified (Figure 200, lane 11) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 250). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1562-

**Example 1416**

A DNA sequence (GBSx1501) was identified in *S.agalactiae* <SEQ ID 4349> which encodes the amino acid sequence <SEQ ID 4350>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL   Likelihood = -11.73   Transmembrane   20 - 36 ( 14 - 42)
    INTEGRAL   Likelihood = -5.20   Transmembrane   207 - 223 ( 206 - 228)
    INTEGRAL   Likelihood = -4.35   Transmembrane   78 - 94 ( 73 - 97)
    INTEGRAL   Likelihood = -4.09   Transmembrane   113 - 129 ( 113 - 133)
10  INTEGRAL   Likelihood = -2.39   Transmembrane   174 - 190 ( 174 - 190)

----- Final Results -----
    bacterial membrane --- Certainty=0.5692 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GF:BAA23750 GB:AB009314 proton-translocating ATPase, a subunit
    [Streptococcus bovis]
    Identities = 149/238 (62%), Positives = 180/238 (75%)
20  Query: 1  MESTSNPVSPFLGIDFDLTILAMSLTITITIFILVFWASRKMTIKPKGKNVLEYVYELN 60
    ME+ NPT  GI+FDLTILAMSLT+ I F ++FWA+RKMT+KPKGKN +KYVE V
    Sbjct: 1  NETSVNPTAKVPGIEFDLTILAMSLTITITIFILVFWATRKMTLKPKGKNFYIEYEFV 60
25  Query: 61  NNTISQNLGHYTKYSLMFLPSFVFIAANLGLMTSLKTHEHFWTSPTANFQVDITLS 120
    NTI NLG YT YSLMF F P+ LNNLGL+ L++ ++NFWTSPT+ VD T S
    Sbjct: 61  QNTIKPNLGEYTPKYSLMFTFFPFILANLGLLVLESEDFNFWTSPTSTIMVQCTNS 120
30  Query: 121  LWAFTCHIEIRKKIGGVLYKGLSPPTPAMLMOLLLEEVNVALELFGNIFSGEVV 180
    L+VA + H+RG+RKKG+ YLKG+LSP P MLPM+L+E+ TNV SLALRFGNT++GSEVV
    Sbjct: 121  LIVAIVHVHVRKKGKVKATLKOTLSPFPMMLPMILEQFTNVLSLALRFGNTYAGEVV 180
35  Query: 181  TGLLLQLAVLSPPTGFLAFALNIWTFAPSMFIGFIQAYVFILSSSYIGHKVDDEE 238
    T L++ S P A ALN+ W AFS FIG IQAYVF ILSS YI K+ DE+H
    Sbjct: 181  TALIVGFGTKSLIPAPFALALNLAWVAFSAFIGICIQAYVFITLSSKYISEKLPEDEE 238

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4351> which encodes the amino acid sequence <SEQ ID 4352>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 33
    >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL   Likelihood = -4.73   Transmembrane   79 - 95 ( 72 - 97)
    INTEGRAL   Likelihood = -4.35   Transmembrane   115 - 131 ( 112 - 132)
    INTEGRAL   Likelihood = -2.13   Transmembrane   200 - 216 ( 197 - 216)
45  ----- Final Results -----
    bacterial membrane --- Certainty=0.2890 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 124/239 (51%), Positives = 169/239 (69%), Gaps = 3/239 (1%)
55  Query: 1  MESTSNPVSPFLGIDFDLTILAMSLTITITIFILVFWASRKMTIKPKGKNVLEYVYELN 60
    ME P + I F+LT+LA+ ++TI I+ F VFWASR+M ++P+GKQ LEX+ V
    Sbjct: 1  MEEAKIPMLKLGPTITNLTILAVCIIVTILVFAVPFWASRQMLKLPKGGKQALYELISFV 60
    Query: 61  NNTISQNLGH-YTKYSLMFLPSFVFIAANLGLMTSLK-T-EHFNFWTSPTANFQVDIT 118
    ++L H K+YSLF+ F+ FV +ANNLGL T L+T + +N WTSPTAN D+
60  Sbjct: 61  DCIGEEHLDNLKQSYSLLLFTIFLPVAVANNLGLFTKLETVNGVNLWTSPTANLAFDLA 120

```

Query: 119 LSLVAFHIGKIRKKGIGYGLKFLSTPDMPLMLLESLVTSVALAHLFQNFQSR 178  
LSL + + HIGKRA + + KLA + + PDM PMILLR TL SLAHLRFQNFQSR  
Sbjct: 121 LSLFTLMVHIGVRRRGLVAHLKLTAPW-MTPMLLEEFTNFLSLAHLRFQNFQSR 179

Query: 179 VTIGLLQLAVLSPTFGPLAFLNIVTAPSMFICIQAYFVILSSSYIGHKHVGDSE 237  
VTIGL+QLA + +AP + + MTAPSF+ IQAYFV L++Y+GK V+ EE  
Sbjct: 180 VTIGLVGLAVLVYVMTAPLFLNIMMTAFVSFVICISQAFVFTKLTYGKGVSE 238

Lipop: Possible site: -1   Crend: 1  
McG: Discrim Score:       -3.50  
GvH: Signal Score (-7.5): -3.36  
      Possible site: 29

```
>>> Seems to have no N-terminal signal sequence
ALOM program      count: 5 value: -11.73 threshold: 0.0
INTEGRAL          Likelihood = -11.73 Transmembrane      20 - 36 ( 14 - 22)
INTEGRAL          Likelihood = -5.20 Transmembrane      207 - 223 ( 206 - 228)
INTEGRAL          Likelihood = -4.35 Transmembrane      78 - 94 ( 73 - 97)
INTEGRAL          Likelihood = -4.09 Transmembrane      113 - 129 ( 113 - 133)
INTEGRAL          Likelihood = -3.39 Transmembrane      174 - 190 ( 174 - 190)
PERIPHERAL        Likelihood = 5.30 156
modified ALOM     scores: ...2.85
```

----- Final Results -----

```
bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

ORF01818 (301 - 1014 of 1314)  
 GP|2662322|dbj|BAA23750.1||A909314 (1 - 238 of 239) proton-translocating ATPase, a subunit  
 [Streptococcus bovis]  
 %Match = 35.0  
 %Identity = 62.2 %Similarity = 78.6  
 Matches = 148 Mismatches = 51 Conservative Sub.s = 39

```

204      234      264      294      324      354      384      414
XANCQTLMFPGVFIERYFLRSICVYILSKIDCNLEKKKG+GLESTSNPTVSPLGIDFLTILAMSLTTIITIFILVFWA
::|||::|:|||:|||||:|||||::|::|:|
NETSVNPTAHVFGIEFDLTILAMSLTVIISFCIIFWA
                        10        20        30

```

SRKMTIKPKGKNVLEYVYELVNNTISQNLGHYTKNYSLLMIFLPSFVFIAANMLGLMTSLKTHEHFWTISPTANFGVDIT  
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||  
TRKMTLKPKGKNVEFYVFPQNTIKPNLEWTPKYSLIMTFPPFFILIANNGLLVLKSEEDYNFWTSISTTMVDCST  
50           60                 70                 80                 90                 100                 110

[illegible]

```

924      954      984      1014      1044      1074      1104      1134
FALNIVWTAFSMFIGFIQAYVFIISSSYICHKVKHGDEEB*EKRGEIQYLLIVQRILVISLSEYALCFPSYLS*LRLHGN
:|::||| || | || | || | || | || | : |:|:
LALNLAWVAFSAFIGCICQAYVFITLSSKYISEKLPEDEDET
          210          220          230

```

-1564-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1417

A DNA sequence (GBSx1502) was identified in *S. agalactiae* <SEQ ID 4353> which encodes the amino acid sequence <SEQ ID 4354>. This protein is predicted to be ATP synthase c subunit (atpE). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL    Likelihood = -4.62    Transmembrane    48 - 64 ( 42 - 65)

----- Final Results -----
bacterial membrane --- Certainty=0.2848 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23749 GB:AB009314 proton-translocating ATPase, c subunit [Streptococcus bovis]
Identities = 55/65 (86%), Positives = 59/65 (90%)

Query: 1 MNLAILALGFVAVMGVIGEGILVANIAKSAARQPEMFSLQTLMPITGVAFIEGTFFVLFA 60
      +NL ILALG AV+GV8+GGGILVANIAKSAARQPEMFSLQTLMF GVAFIEGTFFVL A
Sbjct: 2 LNLKILALGLAVLGVSLGEGILVANIAKSAARQPEMFSLQTLMPITGVAFIEGTFFVLLA 61

Query: 61 FTFIV 65
      TF V
Sbjct: 62 STFFV 66
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4355> which encodes the amino acid sequence <SEQ ID 4356>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL    Likelihood = -5.26    Transmembrane    47 - 63 ( 41 - 64)

----- Final Results -----
bacterial membrane --- Certainty=0.3102 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00920 GB:AF001955 UncE [Streptococcus sanguinis]
Identities = 50/66 (75%), Positives = 58/66 (87%), Gaps = 1/66 (1%)

Query: 1 MNPIF-ALALACPGVSLAEGFLMANLFKAASRQPEITIGQLRSLILGIAFIEGTFFVLTV 59
      MN F L ACPGV8+AE8 +M+NLFKAA8RQPEITIGQLRSLILG+AP+EGTFFVTL
Sbjct: 1 MNLTFGLGACFACPGVSLAEGFLMNSNLFKAA8RQPEITIGQLRSLILGIAFIEGTFFVTLA 60

Query: 60 MAFILK 65
      MAF++K
Sbjct: 61 MAFVIK 66
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 33/62 (53%), Positives = 45/62 (72%)

Query: 5 ILALGFVAVMGVIGEGILVANIAKSAARQPEMFSLQTLMPITGVAFIEGTFFVLFAFTFLVR 66
      I AL A GVS+ EG L+AN+ K+A+RQPE+ +L++LM GVAFIEGTFFV F++
Sbjct: 4 IFALALACPGVSLAEGFLMANLFKAASRQPEITIGQLRSLILGIAFIEGTFFVTLVMAFILK 65
```

-1565-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1418**

A DNA sequence (GBSx1503) was identified in *S.agalactiae* <SEQ ID 4357> which encodes the amino acid sequence <SEQ ID 4358>. Analysis of this protein sequence reveals the following:

Possible site: 47  
>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2562 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1419**

A DNA sequence (GBSx1504) was identified in *S.agalactiae* <SEQ ID 4359> which encodes the amino acid sequence <SEQ ID 4360>. This protein is predicted to be bacterial glycogen synthase (glgA). Analysis of this protein sequence reveals the following:

Possible site: 16  
>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1574 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAAL9591 GB:D87026 bacterial glycogen synthase [Bacillus  
stearothermophilus]  
Identities = 220/475 (46%), Positives = 312/475 (65%), Gaps = 1/475 (0%)

Query: 1 MKIMFVAECAPFAKTTGGLEWIGALPKSLKSGHGVAVMPYDMVDQKFGDIENLMY 60  
MK+P +E APFAK+GGG DV GALPK L + G D V+P Y+ + ++ ++ +  
Sbjct: 1 MKVLEWSECAPFAKSGGLADVAGALPKELRRLGIDARVMLPKYETIAPWKKKAKKVAR 60

Query: 61 FYTDVGMRRHQYGVKRLSQDNVTFYFIDNQYYFYRGHVYQDMDDGERPAYFQLAALRLME 120  
VGR QY GV+ L D V +YFIDN+YYP R +YG +DDGERPAYF A LE++  
Sbjct: 61 LIVFVGMRHQYQVVKELRHGVIYFYFIDNEYFYKRPOLYGHYDGERPAYFCHAVLEVLP 120

Query: 121 KIDTFDVLIVHDYHTAMI PFLLEKKYHNIQAVNNIRAVFTIHRIEFGQGFPEMLGLF 180  
+I F PDV+H HD+HT M+PFLLE+Y Y ++R VFTIHN++FQG F +L DL  
Sbjct: 121 RIQFQPDVHCHDHTGMVFPFLLEBQYRHELFYVIMRTVFTIHNLQGLFPRGILEDL 180

Query: 181 GVGAERYEDGTLKNNKNCINMMKRAILYSRDTVTSPSYANIKITPEFGKGLQDIMMENG 240  
+ + L + C++MK A++ SD +TIVSP+Y EI+T +G+ LD +D +  
Sbjct: 181 NLGGRYFTVDHLEFYGCVSPMKGLVASDLTITVSPITYKRIQTAYYGRKLDGLLRARD 240

Query: 241 KLSGIVNGILDSOLINPETDAFLPYHFSKNNLEKKIKNLALQNLGLPQKQNVPLIGVVS 300  
L GI+NGID + NPE D FL +S E K NK ALQ GLP+ +VPL +V+  
Sbjct: 241 DLIGLINGIDDEFYNPEADFFLTATYSVHTREKQLNKRALQKGLPQKQNVPLIGVVS 300

-1566-

Query: 301 RLTDQKGFILIASELNMLQQDIQMVILDTGVHHPSETPSYFASRYPKLANITFDLRL 360  
 R+T QKQ D++ M+ +D+Q+V+LGTG FE+ FS A+ YP K+ I F L  
 Sbjct: 301 RMTAQKGLDLVTCVPHHMSSEMQVLVLGTGDRFBQPPSQMAAAYPGKVGVIYGHPEL 360

5 Query: 361 AQQIYAASDIPMPSPAFEPQGLSQMMQMYGSLPLVHEVVGGLKDTVVAFNQFQDGSQTGPS 420  
 A QIYA +D+F++PS FEPQGLSQM+A+RYG++P+V E GGL DTV ++N+ G GPS  
 Sbjct: 361 AHQIYAGADLEFLIPSLFEPQGLSQMLALRYGTITVIRFTQGLMDTVQSYNEITKGNGPS 420

10 Query: 421 FNHFSQYMLMQTLKLALVVVNDYPEAMKLLQWAMEKDFSNOTACVAYVQLYQQL 475  
 F +P+ + ++ T++ AL Y P W++L +AM D+SW + Y+Q Y+QL  
 Sbjct: 421 FTFNFAHMLTYTIRPALSFTYQ-PSVWEQLTERAMRGDYSWRRSANQYKQAYDQL 474

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1420

A DNA sequence (GBSx1505) was identified in *S.galactiae* <SEQ ID 4361> which encodes the amino acid sequence <SEQ ID 4362>. This protein is predicted to be a subunit of ADP-glucose pyrophosphorylase. Analysis of this protein sequence reveals the following:

20 Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.3492 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:BAAL9590 GB:D87026 subunit of ADP-glucose pyrophosphorylase  
 [Bacillus stearothermophilus]  
 Identities = 59/178 (33%), Positives = 111/178 (62%), Gaps = 1/178 (0%)

Query: 37 SAEIYVIDITFWLIRQGESEAQNNEPRKLRFLRLDLVBSSNALAFETTYGLNMISSIKSY 96  
 S E+Y++T L++ ++ +N+ + ++RD + +EY+GY + I S++ Y+  
 35 Sbjct: 157 SLEMYLLETSLLELLDADY-KHNGTYSIVDVIRDYHRSLSICEYSGYAAVIDSVQYTF 215

Query: 97 DANMDMLTPNKFYSLFFSNQKVTYTKVNERATYFDRQSNVNSQLASSIIRGYLDHSIV 156  
 ++M++L + + LF + +YTKVK+E T + ++ NV S +A+G +I+G +++S++  
 40 Sbjct: 216 RSMSELLDRDVEQLFLPSHPITYTKVDPEPTYGREGNVRKSMIANGCVIRGTENSVL 275

Query: 157 SRNCLLEKGRVNVNSIIFPKVKIGEGATTENTIIDKCVKVASGVTLKGSILKPLVLPK 214  
 R+ +KG V NSII K +IG+G ++ IIDK KV GV LKG+ ++P ++ K  
 Sbjct: 276 FRSVKIGKGAIVNRNSIIMQKQIGDQCVLDGVIIDKDAKVERGVVLKGTKEQPPIVRK 333

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1421

50 A DNA sequence (GBSx1506) was identified in *S.galactiae* <SEQ ID 4363> which encodes the amino acid sequence <SEQ ID 4364>. This protein is predicted to be subunit of ADP-glucose pyrophosphorylase (glgC-1). Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have an uncleavable N-term signal seq



-1567-

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9765> which encodes amino acid sequence <SEQ ID 9766> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:RAA19589 GB:D87026 subunit of ADP-glucose pyrophosphorylase  
 [Bacillus stearothermophilus]  
 Identities = 195/352 (55%), Positives = 259/352 (73%)

15 Query: 7 MKNEMALILAGGCGIRLGKLTQSIAPKAVQGGYRIIDFALSNCANSINNNGVITQY 66  
 MK + A++LAGGCG+RL LT +IAKPAV FGG+YRIIDF LSNC NSGI+ VGV+TQY  
 Sbjct: 1 MKKCIAMLLAGGCGSRLSLTTNIAKPAVFGGKRYIIDFTLSNCTNSGIDTVGVLTQY 60

20 Query: 67 QPLELNYTHIGNSSWGLDGDISQVTVLPFYSATBNRMFGTSHAIYQNIYDIRINPEY 126  
 QPL L+++IG GS+W LD + GVTVL FYS + G +W++GT++A+YQNI+YI++ NP+Y  
 Sbjct: 61 QPLLIHSYIGISANDLDRKNGGVTVLFPFYSVSSGVKMYEGTANAVYQNIYIBQYNPDY 120

25 Query: 127 VLILSGDHIYQNYDDMLQTHKCNLASLTVAVDVPLKEASRFGIMNTSDNRIVEFEK 186  
 VL+LSGDHIYK+Y NL H A +T++V++VP +EASRFGIMNT+ IVEF EK  
 Sbjct: 121 VLVLSGDHIYQNDYQHMLDYHIAKQADVTVSVIEVMEASRFGIMNTNEEMIVEFEK 180

30 Query: 187 PZHPESTKASMGIIYFDWKKRLRTVLIDGKNGIDMSDFGNKIPAYLESGERVITYNFDG 246  
 P PES ASMGIIYF+W L+ L N DFGK+VIP L +R + Y F+G  
 Sbjct: 181 PARPKNTLASMGIIYFNWPLKQYLDQIDNANPSSHDFGKDVIMLLREKKRFPAYFPFG 240

35 Query: 247 YKDKVGTIESLWEANNEYIGEDNKLHSRDSRWCIYSQNLAPFNFTMDANVDSLVDG 306  
 YKDKVGT++SLWEANN+ +E+N+L DRSW+YS N PP +++ +A V DSLV +G  
 Sbjct: 241 YKDKVGTYSKLEANNLDDNNELDLDRSWRIYSVNPQPPQYISFEAEVSDSLVNEG 300

Query: 307 CFVAGNVEHSILSTNVQVKPNAIIKDSFVMSGATIGEGAKINRAIIGRDAVI 358  
 C V G VE S+L V++ A++K+S +M GA + EGA + RAI+ D++I  
 Sbjct: 301 CVVEGTVERSVLFGQVRIGKGAUVKESVIMFGAAVSEGAYVERAIVTDSII 352

There is also homology to SEQ ID 2660.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1422

A DNA sequence (GBSx1507) was identified in *S.agalactiae* <SEQ ID 4365> which encodes the amino acid sequence <SEQ ID 4366>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2844 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA76440 GB:Z14057 1,4-alpha-glucan branching enzyme [Bacillus  
 caldolyticus]  
 55 Identities = 272/616 (44%), Positives = 371/616 (60%), Gaps = 14/616 (2%)

Query: 6 ELYTFGIGENFELQNYLVGHSKNGSFC---PRVWAPNAENNVQIGDFDWRNRLPQNK 61  
 E+Y F G + G H G F VWAP+A V++G F DW + K

-1568-

Sbjct: 10 EYVLFPHRGRLYQSYELFGAHVIRGGGAVGTRFCWAPHAREVRLVGSFNDWNGTNSPLTK 69

Query: 62 -NQAGVWEANSILDAREGDLKYLVLTREGGQVVEKIDPMVYMERPPGTASIVKLRNKKM 120  
N GVM + EG DYKY + G+V+ K DP A Y E R P TAS++ L+ +N

5 Sbjct: 70 VNDSEWVTIVVPEHLSGLYKYEIITFDGRVLKADPYAFYSELRPTASIVYDVGKSYEN 129

Query: 121 EDGLAMGRRKRRLGSPQRPINIVSYHAGSNKKDDFGHPMTFSQLDKDYLIPLYVEMSYTHVE 180  
D W +++R +P+ IYS+H GSWKK G T+ ++ D LIPT++K +TH+Q

10 Sbjct: 130 NDSFWQRKRRKRRLGYDQPMVIVSYLHFGSWKKKPDGRPTTYREMADELIPYVLENGPHTIE 189

Query: 181 FMPMLAHPLDMGWCYQLMGYFAFHTYGTTFEFPQDFVSEACHKNNIGLVLDVPGHFQIND 240  
+PL+ HFLD SWSYQ GI++ YGTP +P FV+ CH+ +GV++DWVPGHF ++

15 Sbjct: 190 LLLPLVEHPLDRSGWYGTGYISVTERKGTIPHEFMIFVDRCHQAGLGVIIDNVPGHPCKDA 249

Query: 241 DALAYFDGTATTEYQNHDAIRNRWAGALNFDLGNKQVQSLISSALFNIERYHIDGIRVD 300  
L FDG TTEY N NY WG NFDLKG +V+SFLIS+ALFW+E+YH+DG RVD

20 Sbjct: 250 HGLNFDGAPTYEANSKDRBNVWGTANFDLGRKPEVRSFLISSALFNLEYHYVDGFRVD 309

Query: 301 AVENNLNLYLDYDGGWEANQFGDNRRLLDGHFLRKLAKVIKREHNWMMIAESSTASTPT 360  
AVANGLY ++ +R N FLR-LN+ + FNV MIAE+ST +T

25 Sbjct: 310 AVANGLYWRNDRLYE-----NFAVEFLRLQNEAVPAYDFNVMMIAEDSTDWPRVT 361

Query: 361 KDLESGLQDFDGGNHWNDILRFTPEEPLRYQYDFNLATTSFMYIFNENFLVAFSHDE 420  
GGAP+ +NKGWMD+L++ E P R+Y N V+FS +Y ++ENP+L FSHDE

30 Sbjct: 362 APTYDGLGAPNTQNGWMMNMLKYMETPPHERKYNBQVFSLLYAYSENPIFFSHDE 421

Query: 421 VVHGKKSMMHGWNDRYNQFAGLRHLYAVQWCHPGKLLFWGSEFGOFLNKNYNDQLEWE 480  
VVHGKKS++NM G +FR LR LY Y M HPGKLLFWGSEF QF ENK+ ++L+W

35 Sbjct: 422 VVHGKKSLLRNPGSYBEKPAQLRLLYGYMAHPGKLLFWGSEFQFENKFAELDLYV 481

Query: 481 NNDNDMMNQKRYTKLQALQFYKDHKCLWRIDDSFDGLSIIDANKSETSLVSPIRKDK-G 539  
+ +++NM Y KQL YK +K +D G E ID N +++ SFIR+ K G

40 Sbjct: 482 LPDFELHRMDYVQKLIACYKRYKPFYELDNDHDPFGFENIDVHRAQSIFSIFIRKQKGG 541

Query: 540 DLLLCVPMFPERNPTIGVPGAGIYEEVLNTEMEFPGVWMMNVPVTKCVAIMDYD 599  
D+L+ V N T ++ + VP Y EVLN+ EFGG + +

45 Sbjct: 542 DVLIVCNFTNQDYDQYKVSPLAPYREVLSDAHPGSSGHVNGKRLPAFSEFPFGKP 601

Query: 600 HTLSFTLFAFGASVWR 615  
+ + T+P G S+ R

50 Sbjct: 602 YEVMTTIPFPGISILR 617

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 45 Example 1423

A DNA sequence (GBSx1508) was identified in *S.agalactiae* <SEQ ID 4367> which encodes the amino acid sequence <SEQ ID 4368>. This protein is predicted to be pullulanase (pulA). Analysis of this protein sequence reveals the following:

Possible site: 45  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3194 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:RAC4685 GB:U67061 pullulanase [Bacteroides thetaiotaomicron]  
Identities = 223/597 (37%), Positives = 331/597 (55%), Gaps = 55/597 (9%)

60 Query: 139 EYSETKTAFLWAPTAERVELLYHSTDETASVSKVLSMKGTAVNYKNKENTHGVWFT 198

-1569-

```

      EY+   T F LW+PTA+ V L+LY +   E   + + M+ G           G W
Sbjct: 46 EYTPKATKFTLMSPTADEVRMLLYEA-GEQGHAYTIVKMSGE-----EGTWTA 93

5  Query: 199 ELEGNHNYQAYTYRVYVYRRRTFKITRDPSYIATANGKRSIVLAPALTPGFKSHKSGE 258
      + + + YT+ V + + T + A   NGKR+ +L + + P+G++ + +
Sbjct: 94 VVSKDLIGKPYTFHVKLDDKQKQDTPGINARAVGVNGKRAALIDWQSNPDGWE-----SD 149

      AKWRLENPNQAVIYEMHRDPSISSETSGVKIDYHGKFKGLHQKGTVNQHGDKTTDFYVDQ 318
      + L++P + IYEMH RDFS+ TSGVK   GK+ L + GT+N   T D++ +
10  Sbjct: 150 TRPPLKSPADMIYIEMHRDPSVDSSTSGVKNK--GKYLALTEHGTVNSDKLLTGIDHLIE 207

      LGVNYQLQPIFDHHCQTFDD--GHAYNNGSDPENYNVPEASFSNNEPATRILELKSJA 377
      LGV ++ L P FD+ + + YNNGSDP+NYNVP+ S++++P++PATR+ E K
Sbjct: 208 LGVTHVHLLPSFDIASVDVETRLNNSYNWGTDPQNYNVPDGSYATDPYQATRVKRFQM 267

15  Query: 378 IQAYHDAGIGVIMDVVYNNHTPSSDSAPQLTVZDYYYNNMNHGTFQWSSCGNETASEKE 437
      +QA H AGI VIMDVVYNNHTF++ +S P+ TYP Y+YR + T   NGSOGNETASS+
Sbjct: 268 VQALHAGIRVIMDVVYNNHTFDSINFERTVPGYFYRQKEDKTLANGSOGNETASERL 327

20  Query: 438 MCKRYILDSVLVWKEYNIDGPRFDLMGLHDVYKTMNIRNELNKIDPRIAYGSDWMDGA 497
      M RK+++SVLYW+KEY++DGPRFDLMG+HD+ETVN IR +N +DP I +YSGW A
Sbjct: 328 MMRKFWVESVLWIKETHVDGPRFDLMGHDHETVNEIRKAVNAVDPTCIYSGWMAARA 387

25  Query: 498 GLTPQNK-AKKDANQMPGIGFREDVRDAV---KGAEIYGFKGLVNSGNETDIVAKG 553
      F + A K N Q+PG+ F+D++RD + G + G F G+ G E V G
Sbjct: 388 PQYPADSLAMKGNIAIQPVAVFSGDELRDGLGFPVSGDKKGAFLAPGIPGG---EMSVKFG 444

      ILGSDS-----LVSYL-----DPSQVLYTVYEAHDNYNLDLMLHFNDEKHIYR 600
      I G+ E   V+Y   P Q++YV HD L D L P+ +Q I
30  Sbjct: 445 IAGAIEHPQVQCDSDVNYTKPWAKQFVQMIISYVSCDGLCLVDRLLKASMDITFELQIRL 560

      VEVASAMLLAQGMAMQGLQGFPLRTKCTPTGDKQGLTQADKERAMNSYAPDQVNVQW 604
      ++A A+   QG+ F+ G+E +R   DK+ NSY +ED VN ++N
35  Sbjct: 505 DKLAQAVVFTSQGIPPIYAGEEIMR-----DKQGVINSYKSPDVANIDW 549

      DNVTFHKSTINFIRKILITKNSFYPYSYSSFEIRKHVPEVESAQYSGFISPTVEEH 717
      T   + + + + L+ + P F   +RKH+ + S I+E + + +
Sbjct: 550 RRKTTSDADVPMYKRLIDLKSHPAFPMGDAGQVRKHLEFLPVE-GSNLIAFLKDH 605

```

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1424

A DNA sequence (GBSx1509) was identified in *S. agalactiae* <SEQ ID 4369> which encodes the amino acid sequence <SEQ ID 4370>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 40
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
50  bacterial cytoplasm --- Certainty=0.2368 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

55  >GP:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
    Identities = 151/293 (51%), Positives = 193/293 (65%), Gaps = 5/293 (1%)

      KKARLIYNFTSGQEIIMKKNVAKVLIDLEQPGYKTSAPQTTPTKNSARDENETPAACQAGFDL 64
      K+AR+IYNFTSG+EI KK+AA+VL E GYKTS TT A A AA FDL
Sbjct: 2 KRARIYNFTSGRRIFKHLAQVLQKFEQAGYKTSHTATT-CAGDAVHAQKAALREFDL 60

60  Query: 65 IVAAGGDTGTFINEVNGIAPLIGRRPKMALIPTGTINDFARALKIPRGNIPKATKLIKGNQI 124
      I+AAAGGDTGTFINEVNG+APL RP + +IP GTINDFARAL IPR + +A +

```

-1570-

- 5  
10  
15  
20  
25  
30  
35  
40  
45  
50
- Sbjct: 61 IIAAGGDTINEVVGGLAPLDRPTLGVIPVGTINDFARALGIPRSDLKAADTVINGVA 120  
 Query: 125 VKMDIGQAQEDNYFINIAAAGSLTSLTYSVPSQLKITPGYLAYLAKGVELLPRVRKVPVK 184  
 +DIGG YFINIA G LIRITY VTS+LET G LAY KG+R+LP +R V+  
 Sbjct: 121 EPDIDGQVN-GQYFINTAGGGRILTLTYDVPSKLTMLGQLAYYLKGMPLSLRPTVEVR 179  
 Query: 185 ITHDKGEFIGDASMIFFVAITNSVGGFQIAPDAKLDDGKFTLILVKTANILKIMHLIRIV 244  
 I +D F G+ + V +TNSVGGF+ +AD+ L+DG F L++K ANL E + + +  
 Sbjct: 180 IETDGKLPQGEIMLFLVTLTNSVGGFQIAPDAKLDDGKFTLILVKTANILKIMHLIRIV 239  
 Query: 245 LAGGKHINDKRVVEYIKTSYLTIEPLSDERMMINLDCGYGGDAPITLANLKHNI 297  
 L G+HIND+ + Y K + + + B+M +HLDGSGOG P NL HI  
 Sbjct: 240 LR-GEHINDQHIIYTKANRVKNV--VSRKMQLMDGEGYGGDLPGEFVNLVRI 289  
 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4371> which encodes the amino acid sequence <SEQ ID 4372>. Analysis of this protein sequence reveals the following:  
 Possible site: 40  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2501(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 An alignment of the GAS and GBS proteins is shown below.  
 Identities = 272/334 (81%), Positives = 300/334 (89%)  
 Query: 1 MKKQKARLIYNPTSGQEMKNQVAEVLIDBGGFYETSAPQTTFKNSGARDEATRAAQA 60  
 MKKQ +ARLIYNPTSGQEM+K+V EYLEILBGGFYETSAPQT KNSA +EA RAA+A  
 Sbjct: 1 MKKQARLIYNPTSGQEMKNQVAEVLIDBGGFYETSAPQTTFKNSGARDEATRAAQA 60  
 Query: 61 GFDLVAAGGDTINEVVGGLAPLDRPTLGVIPVGTINDFARALGIPRSDLKAADTVINGVA 120  
 GFDL+VAAGGDTINEVVGGLAPLDRPTLGVIPVGTINDFARALGIPRSDLKAADTVINGVA  
 Sbjct: 61 GFDLVAAGGDTINEVVGGLAPLDRPTLGVIPVGTINDFARALGIPRSDLKAADTVINGVA 120  
 Query: 121 KNOIVMDIGQAQEDNYFINIAAAGSLTSLTYSVPSQLKITPGYLAYLAKGVELLPRVRK 180  
 KNO +MDIG+R+D YFINIAAAGSLTSLTYSVPSQLKIT PGYLAYLAKGVELLPRV  
 Sbjct: 121 KNOIQMDIGRAKDTYFINIAAAGSLTSLTYSVPSQLKITPGYLAYLAKGVELLPRV 180  
 Query: 181 VPVKITHDKGEFIGDASMIFFVAITNSVGGFQIAPDAKLDDGKFTLILVKTANILKIMHLIRIV 240  
 VPVKITHDKG F G SMIF AITNSVGGFQ IAPDAKLDDG FTLLI+KIMHL EI+HL  
 Sbjct: 181 VPVKITHDKGVFBQGVSNIFPAITNSVGGFQIAPDAKLDDGFTLLI+KIMHLIRIV 240  
 Query: 241 IRLVLAGGKHINDKRVVEYIKTSYLTIEPLSDERMMINLDCGYGGDAPITLANLKHNI 300  
 +RL+L GGGKH D+RVYKIKTS + IEP +RMMINLDCGYGGDAPITI NLKNI PF  
 Sbjct: 241 LRLVLAGGKHINDKRVVEYIKTSYLTIEPLSDERMMINLDCGYGGDAPITLANLKHNI 300  
 Query: 301 ANTDEISDQALWLDLMDLALHIAQKFAVEVDL 334  
 A+TD ISDQALWLDL I R I +KFA+EV+DL  
 Sbjct: 301 ADTLDISDQALWLDLMDLALHIAQKFAVEVDL 334

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1425

- 55 A DNA sequence (GBSx1510) was identified in *S.agalactiae* <SEQ ID 4373> which encodes the amino acid sequence <SEQ ID 4374>. This protein is predicted to be DNA ligase (fig-A-1). Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence

-1571-

INTEGRAL Likelihood = -0.27 Transmembrane 363 - 379 ( 363 - 379)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.1107 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9763> which encodes amino acid sequence <SEQ ID 9764> was also identified.

- 10 The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB12482 GB: Z99107 similar to DNA ligase [Bacillus subtilis]  
Identities = 346/657 (52%), Positives = 462/657 (69%), Gaps = 8/657 (1%)

- 15 Query: 2 ENRMNSLVSLNQAKEYTQUNPTVSDSQYDQLYRELVELSKQHPENTLPMSPTRVGG 61  
+ R EL +N+Y+ EYTT D P+V D++YD+L +EL+ +E++HP+ P+SPT RVGG  
Sbjct: 7 KQRAELRRTINKYSTETTYLDEPSVPOAEYDRLMQELIAIBEEHFDLRTPOSPTORVGG 66
- 20 Query: 62 LVLGFEKYQHSEYPLYSQAPFSKEELIAFQKRVKAEF-PTAAYMAELKIDGLSVSLTYV 120  
VLE P+K H P+ SL +AP+ ++L FD+RV+ AY ELKIDGL+VSL Y  
Sbjct: 67 AVLEAPQKVTHGTPLMSLGNAFNADDLRDFORVRQSGVDVAYNVLEKIDGLAVSLRYE 126
- 25 Query: 121 NGVLQVGATRGDGHIGENITENLKRVDIPLHLQSLDITVRGECYLPKESFEAINTEKR 180  
+G GATRGDG GE+ITENLK + +IPL +++ L I VRGE Y+PK SFEA+N E+  
Sbjct: 127 DGYFVRGATRGDGTGTDITENLKTIRNIPLMQNRLESLVGRGAIMPMSFEALNTERI 186
- 30 Query: 181 ANGEQEFANPRNAAAGTIRQLNTGIVAKRKLATPLVQESPTQK--ETQDDVLEKESYG 238  
N E+ FANPRNAAAG+LRLQ+ I AKR L P+Y A + FTQ L L+ G  
Sbjct: 187 KNSEFPFANPRNAAAGSLRLDPKIAAKRNLDPVYSIAELDEMGYSQGLDPLBELG 246
- 35 Query: 239 FSNVHRRLISSMEKIVDPFTQIEKDRVSLPYDINGIVIKVNSIAMQELGFTVKAPFWA 298  
F N R S+E++ I ++ R LPY+IDGIVIKV+G+ QESLFT K+KFWA  
Sbjct: 247 FKTQQRKQKSIEEVITIDELQAKRADLPYIDGIVIKVDSLDQSELGFTKAPFWA 306
- 40 Query: 299 IAYKFPABEKEKILSVNIVGRTGVVTPITANLTPVQLAGTIVSRATLHNVDTAEKDIR 358  
IAYKFPABE ++L ++ VORTGV+TPTA L PV++AGTIVSRA+LHN D I EKDIR  
Sbjct: 307 IAYKFPABEVVTKLLDIELNVGRTGVITPTAILEPVVAGTIVSRASLHNEDLKEKDIR 366
- 45 Query: 359 IGDTVVVYKAGDIIPAVLANVMKRNQCEVML-IPKLCPSCGSELVHPGEVALRCINPL 417  
I D VVV KAGDIIP V+N++ +R +E +P CP CGSELV RGEVALRCINP  
Sbjct: 367 ILDKVVVKAGDIIPVNVVLVDQRTGEEKFSMPTECPGSELVRIBGEVALRCINPE 426
- 50 Query: 418 CPWQIKERLAHFASRDAMNITPGPSPVLVEKLFDAHLIADVADLYSRITRNLAITDGIKEK 477  
CP QI+E L HP SR+AMNI G G ++ +LF+ +L+ +VAD+Y+L+ E ++ L+ + EK  
Sbjct: 427 CPAQIREGLIHFVSRKAMNIDGLGERVITQLFEENLVRRVADLYLTKERVITQERMGEK 486
- 55 Query: 478 SATKVIYHAIQSSKENSARKLLPGLGIRHVGSKASLLLEFPNLRQLSQASQESIASIDG 537  
S + +IQ SKENS E+LLPGLGIR +GSKA+ L P +L L +A+S+E + +D  
Sbjct: 487 STENLISSIQSKENSLERLLPGLGIRPIGSKAAKTAMHPESLENLKKASKEELLAVDS 546
- Query: 538 LGGVIAKSLHTFFPKKEVDLKKRLTSYNNVFNLYG---KRVTDAQLSGITIVLTQKGL 593  
+G +A ++ T+P KEE+ +LL EL VN Y G K +D+ +G T+VLTKGL  
Sbjct: 547 IGEXMADAVTTPYHKKEMLKLLNELQELGVNTLYGPKKVKARSDSYFAGKTIVLTQKGL 606
- Query: 594 EKKMTNEAKKIKNLQKAGKVTSVSKTDLIVAGSDAGSKLTKAQDLGITIQDEIMLL 650  
E++-RNEAK ++ LQ K+TGSVSK TDL+AG AGSKLTKAQ+L I + +E L+  
Sbjct: 607 EKLNRNEAKAQIRALGKLTGSVSKNTDLIVAGEAGSKLTKAQELNIVNHRPQIM 663

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4375> which encodes the amino acid sequence <SEQ ID 4376>. Analysis of this protein sequence reveals the following:

- 60 Possible site: 61  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -0.43 Transmembrane 363 - 379 ( 363 - 379)

-1572-

----- Final Results -----

bacterial membrane --- Certainty=0.1171 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 472/652 (72%), Positives = 556/652 (84%)

Query: 1 MENTRMNELVSLNQYAKGYTQDNPTVSDSDQLYRELVELEKQHPENILPNSPTRVG 60  
 M+ R+ EL LNN+Y +YTT+D P+VSDS YD+LYRELV LE+ +FE +L +SPT +VG  
 Sbjet: 1 MKKR1KELTDLNRYKYDYTTKDAPSVSDSDYDKLYRELVLTSQSYPEYVLQDSPTQVG 60

Query: 61 GLVLEGFPEKYQHEFYPLSLQDAFSEKELIAPDKRVKARFPTAAYMAELKIDGLS+SL+YV 120  
 G +L+GFPEKY+H+YPL+SLQDAFS+EEL APDKRVKARFP A Y+AEKIDGLS+SL+Y  
 Sbjet: 61 GTILKGFPEKYRHQYPLSLQDAFSEKELIAPDKRVKARFPNATYLAELKIDGLSISLSYE 120

Query: 121 NGVLQVGATRGDGNIGENITENLKRVDIPLHLDQSLDITVRGECYLPKESFEAINIEKR 180  
 NG LQVGATRGDGNIGENITEN+K++ DIP L + L ITVRGE Y+ ++SF+AIN ++  
 Sbjet: 121 NGFLQVGATRGDGNIGENITENIKKIDIPYQLSERPLITTVRGEAYMSQSFKAINEARQ 180

Query: 181 ANGEQEFANPRNAAGTLRQLATGIIVAKRLATFLYQASPTQKETQDDVLKELSES+YGS 240  
 NGE EFANPRNAAGTLRQL+T +VAKR+LATFLYQASPT + Q++VL EL GSS  
 Sbjet: 181 ENGETEFANPRNAAGTLRQLQTSVAVKRLATFLYQASPTARNQNEVLAEALDGLS 240

Query: 241 VNHRLISSNRKIWDFIQTIEKDRVSLFYDIDGIVTKVNSIAMQRELGFTVKAPRMAIA 300  
 VN + ++SS++IWDFI+TIE R L YDIDG+VTKVNS+AMQRELGFTVKAPRMAIA  
 Sbjet: 241 VNFYYQLTSSNDEIWDFIKTIRAKRDQLAYDIDGVTKVNSLAMQRELGFTVKAPRMAIA 300

Query: 301 YKFAEKEKAEILSDVWTVGRTGVWTPANLTPVQLAGTTVSRTATLENVDYIAEKDIRIG 360  
 YKFAEKEKAEILSDVWTVGRTGVWTPANLTPVQLAGTTVSRTATLENVDYIAEKDIRIG  
 Sbjet: 301 YKFAEKEKAEILSDVWTVGRTGVWTPANLTPVQLAGTTVSRTATLENVDYIAEKDIRIG 360

Query: 361 DTVVYKAGDIIPAVLNVMSKRNQGEVWMLPKLPCSCGSELVHPFEDEVALRCINPLCPN 420  
 DTV-VYKAGDIIPAVLNVMSKRNQGEVWMLPKLPCSCGSELVHFE EVALRCINPLCP+  
 Sbjet: 361 DTV-VYKAGDIIPAVLNVMSKRNQGEVWMLPKLPCSCGSELVHPFEDEVALRCINPLCPN 420

Query: 421 QIKERLAHFASRDAMNITGPGPSLVEKLFDAHLIADVADIYRLSIENLLTLDGIKEKSA 480  
 I+ L HFASRDAMNITG GP++VEKLF A + DVADIY+L+ E+ + LOGIKEKSA  
 Sbjet: 421 LIQRSLHFASRDAMNITGLGPAIVEKLFAGFVHDVADIYQLTDFMQLDGIKEKSA 480

Query: 481 KIYHAIQSKSENSAKILFGLGIRHVGSKASRLLEEFBNRLQSLQASQESIASIDGLG 540  
 K+ AI++SK NSAKILFGLGIRH+GSK SRL+LE +G++ L A +E IA IDGLG  
 Sbjet: 481 KLLAAIEPASKSENSAKILFGLGIRHVGSKVSRLLIEVVDISALLTAKEEETARIDGLG 540

Query: 541 VIAKSLATPFKEEVDKLLLELTSYNFMNVLGKRVSTDAQLSGTLVLTGKLEKMTNRE 600  
 IA+SL +FE++ L++EL + VN +Y G++V++DA L GLTVLTGKL ++ RNE  
 Sbjet: 541 TIAQSLTQYFQKTAAILVDELTAGVNMHYSGQKVNDAALGLTVLTGKLANLRNE 600

Query: 601 AKEKQLNLGAKVTGVSVKTTDLIVAGSDAGSKLTKAQDLGITQIDREDWLLNL 652  
 AK+KL+ LGAKVTGVSVKTTDL+AGSDAGSKL KM+ LGI I+DEDWL L  
 Sbjet: 601 AKDKLEALGAKVTGVSVKTTDLIVAGSDAGSKLEKAKSLGIRIREDWLRQL 652

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 55 Example 1426

A DNA sequence (GBSx1511) was identified in *S.agalactiae* <SEQ ID 4377> which encodes the amino acid sequence <SEQ ID 4378>. Analysis of this protein sequence reveals the following:

Possible site: 32

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

60 INTEGRAL Likelihood = -5.63 Transmembrane 110 - 126 ( 108 - 128)  
 INTEGRAL Likelihood = -2.13 Transmembrane 142 - 158 ( 141 - 159)  
 INTEGRAL Likelihood = -1.12 Transmembrane 75 - 91 ( 75 - 93)

-1573-

----- Final Results -----

bacterial membrane --- Certainty=0.3251 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA68244 GB:X99978 citrulline cluster-linked gene [Lactobacillus  
 plantarum]

Identities = 56/158 (35%), Positives = 91/158 (57%), Gaps = 8/158 (5%)

Query: 13 AIVTAIYIVLITPPFNALAYGAYQFRVSEMLNLAFLYHRKYLEFAVTLGQNLNLYSPG- 71

A+V A+Y+VL + P ++A GA QFRVSE LN LA ++RKY++ + G ++ + G

Sbjct: 13 ALVAMTYVVLICGPAAFSIASGAIQFRVSEGLNHLAVFNKRYINGIVAGVILFDFAGPGA 72

Query: 72 -MIDVFGGSGSTLLFVYLITLILFKQYQKDYLPGLINKAKFFFFSFFFAASMITVAVELKI 130

++V GGG +LL + + T L + K L+N A F S F A MIT+ +

Sbjct: 73 .SLNLVFGGGQSLLALLVLTWLAPKL-KTVQRMFLNLAFTVEMFMIALMITM-----M 126

Query: 131 VAGLEPLLLTWLTAVGELASLLGAVLVNKLSEHVDFT 168

+G+ T+LTTA+ EL + + A + L R + F+

Sbjct: 127 SSGVAFNPTYLITALSELIIMSITAPIMYSLEVDLHFS 164

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4379> which encodes the amino acid sequence <SEQ ID 4380>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -4.41	Transmembrane	75 - 91 ( 70 - 94)
INTEGRAL	Likelihood = -3.82	Transmembrane	12 - 28 ( 8 - 28)
INTEGRAL	Likelihood = -2.28	Transmembrane	141 - 157 ( 140 - 158)
INTEGRAL	Likelihood = -0.64	Transmembrane	110 - 126 ( 110 - 126)
INTEGRAL	Likelihood = -0.59	Transmembrane	55 - 71 ( 54 - 73)

----- Final Results -----

bacterial membrane --- Certainty=0.2763 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 114/167 (68%), Positives = 137/167 (81%), Gaps = 1/167 (0%)

Query: 1 MNTFTTROYARNAIVTAIYIVLITPPFNALAYGAYQFRVSEMLNLAFLYHRKYLEFAVTL 60

M T DY H+ +V A+Y+VLITTFP NAI+YG YQFR+SEM+NLAFLYHRKY+ AVTL

Sbjct: 1 MTKLTVHDYVHIGLAALYVVLITTPFNALAYGAYQFRVSEMLNLAFLYHRKYLEFAVTL 60

Query: 61 GCMINLNSYFGMIDVFGGSGSTLLFVYLITLILFKQYQKDYLPGLINKAKFFFFSFFFAAS 120

GCM+ N YSPG+ IDVFGGSGSTL+ FV LG ILP +YQKDYLPNG+ NKAF +FSFFFA S

Sbjct: 61 GCMINLFYSFGLIDVFGGSGSTLLFVYLITLILFKQYQKDYLPGLINKAKFYFSSFFATS 120

Query: 121 MITVAVELKIVAGLEPLLLTWLTAVGELASLLGAVLVNKLSEHVDFT 167

M VA+EL G P LITW TTA+GEL SLL+G+++DKIS+ + F

Sbjct: 121 MFVAELKLYFF-GAPFLLTWFTTALGELVELLIGSLIDISQKISF 166

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1427

A DNA sequence (GBSx1513) was identified in *S.agalactiae* <SEQ ID 4381> which encodes the amino acid sequence <SEQ ID 4382>. Analysis of this protein sequence reveals the following:

-1574-

Possible site: 53

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

5 INTEGRAL Likelihood =-11.20 Transmembrane 255 - 271 ( 245 - 281)  
 INTEGRAL Likelihood =-10.72 Transmembrane 141 - 157 ( 132 - 165)  
 INTEGRAL Likelihood = -8.17 Transmembrane 189 - 205 ( 185 - 208)  
 INTEGRAL Likelihood = -7.01 Transmembrane 36 - 52 ( 33 - 60)

----- Final Results -----

10 bacterial membrane --- Certainty=0.5479(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AA35915 GB:AF071085 Orfde2 [Enterococcus faecalis]  
 Identities = 83/276 (30%), Positives = 157/276 (56%), Gaps = 3/276 (1%)

Query: 17 RPIQVMRHPOGAEMDLGAIAVAYLLVTAFLVLVIAANIFFYPHINVDLSLMQKRLP 76  
 R I+ H +AE+ S++ VAYLL+ FPL+ N+ PY I+ +L + + P  
 20 Sbjct: 15 RFIETTQSHVMTABIGNSSVVVAYLLSLFLLIAGNVLPYLRIIDPNSVLPYIAEAP 74

Query: 77 KNIVFPASRLAVDAFSPKSTGILGFASLTAFWIMSKSLTSLQKAINKAYGVQHRDFVIS 136  
 K++Y+ ++ S G+L ++L APW+ S+S+ +LQ A+NKA+GV+Q ++P++  
 25 Sbjct: 75 KDVVYNLEPAIRSLTQRSGGLLSVSAALAFWSAQGINALQNMNKAQGVGQRQNFILV 134

Query: 137 RLVGVGTLIIILFLITFVLIFSTFSPKVLQIIVNMYDLGTLTAMLLNLQAFVFTLITFL 196  
 R+V L+ + + V++ + +++++ ++ ++ L P+T + + +  
 30 Sbjct: 135 RVVSVNLVLLFMVAIVGVVILGLSQYIIELQLPIPHYSTVIDTFQALKVPLTIVVLV 194

Query: 197 GIGILYFILPNARIKRVYVIGTLSTFVIGFSPNLSIQYVLNKEVMKDITKPSGVI 256  
 + +Y + +FN ++ +R ++PG +FST S + YV ++ + GS +  
 35 Sbjct: 195 IMCLIVAVPNRKL-SLRSLPGAIFSTVGWMLSQIPGLVYKYPSSRIASYQIGSF- 252

Query: 257 FILMLWPIFLAHIMILGAILNASVQRIATQKIBRR 292  
 IIMLN F A I+LGA+NA V E G E ++  
 35 Sbjct: 253 -IIMLNLPAATIIILGAIVNAVVDVLYLXKKERKQ 287

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4383> which encodes the amino acid sequence <SEQ ID 4384>. Analysis of this protein sequence reveals the following:

Possible site: 53

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood =-12.58 Transmembrane 141 - 157 ( 132 - 168)  
 INTEGRAL Likelihood =-12.15 Transmembrane 189 - 205 ( 177 - 210)  
 INTEGRAL Likelihood =-11.68 Transmembrane 256 - 272 ( 245 - 280)  
 45 INTEGRAL Likelihood = -7.54 Transmembrane 36 - 52 ( 33 - 60)

----- Final Results -----

50 bacterial membrane --- Certainty=0.6031(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA58244 GB:X99978 citrulline cluster-linked gene [Lactobacillus  
 plantarum]  
 55 Identities = 53/170 (31%), Positives = 92/170 (53%), Gaps = 11/170 (6%)

Query: 1 MTKLVVHDVYHIGLVAALYVVLITTPFLNLSYGYQYPTISMENFLAPYHRYKIIAVTL 50  
 MT+ + ++ LVA+YVVL + P +++ G QPR+SR +N LA ++RKYI +  
 Sbjct: 1 MTQSKIRPWITINVAAMVYVVLCLGPAFSLASGAIQPRVSEGLNHLVAFRKRYIWIIVA 50

60 Query: 61 GCMIANFYSFG--LIDFVFGGGSLTIFVTLGVILFSEKYQDYLPNGFIPKNAFYSPFFFA 118  
 G ++ + + G L+V GGG +L+ + + L K + ++ + + + F  
 Sbjct: 61 GVILFDAPFGGASLNVLPFGGQSLLALLVLTWLAFLKLT-----VWQRLNLIA-LFT 113

Query: 119 TSMFVNA--ILEYFFGAFFLLTWFTALGELVSLIGSLIIDKLSQRIQF 166



-1575-

SNP +A I + G F T+ TTA EL+ + I + I+ L + + F  
 Sbjct: 114 VSMFMIALMITMSSGVAFKPYTLTALSELIIMSTIPAFIMYSLDRVLHF 163  
 !GB:AF071085 Orfde2 [Enterococcus faecalis] 176 2e-43

5 >GP:AAC35915 GB:AF071085 Orfde2 [Enterococcus faecalis]  
 Identities = 90/271 (33%), Positives = 155/271 (56%), Gaps = 3/271 (1%)

Query: 19 IQVPMRHLQSAEMDLASIAVAYYLITAFPLIVIAANIPFYINIDIADLLRMKQNLPKD 78  
 I+ H+ +AE+ S++ VAYYL+L+ PFL++ N+ PFL ID +L + + +PKD  
 10 Sbjct: 17 IETQSHMVTARIGNSSVVVAYYLLSLPFLLIAGVGNLPLYLRIDPNSVLPYIARAI PKD 76

Query: 79 IFRPASAIVENIPSKPSGSLGVATLTGLWIMSRSLTSIQKAINKAYGASQHRDFFI GHL 138  
 +++ + + + + SG +L V+ L W+ S+S+ +IQ A+NKA+G Q + +F + +  
 15 Sbjct: 77 VYKNLEPAIRSLTLQRSGELLSVSAALAFWAGSAGSINALQAMNKA PGVGRKNFPLVRV 136

Query: 139 VGLLTSLIILFLAFALIPISPSKAAIQVLDKHYHLSDNITTI FLLLIQPTVLIIFVGL 198  
 V L L+ + + + + + I++L +H S++ F L P+T +++ V +  
 20 Sbjct: 137 VSPVLILFMVAIVGVVVILGLQYIIELOQPIHFYSTSVIDTFAQKMWPLTIVVLVIM 196

Query: 199 NLLYTLLENKVKIKIRYILPOTLFTSFVMTLGNLVGVNVVVERMDIKHFGSVWIFI 258  
 L+Y ++FN K+ +R ILPG +F++ LS + G IV Y R+ ++ GS I  
 25 Sbjct: 197 CLTYAVVENRKL-SLRSLPGRAIPSTGVWMLLSQIFGLYVKYFSSRIASVQIIGS--FII 253

Query: 259 INLWPIFLARILILGAI FNATYQMSLQKE 289  
 +NLW P A T+ILGAI NA E G E  
 30 Sbjct: 254 LNLNLNFAATITIIIGAI VNAVVDYLYKCKE 284

An alignment of the GAS and GBS proteins is shown below.

Identities = 188/302 (62%), Positives = 244/302 (80%)

30 Query: 1 NKLKFFEDLLAKLEYRPIQVPMRHPQSAEMDLASIAVAYYLLVITAFPLIVIAANIPFPY 60  
 K KK+P+ +L+K +Y FIQVPMRQ QSAEMDLASIAVAYTL++TAFPL+VIAANIPFY  
 Sbjct: 1 KAEKQVDFKVLKSKQYEFIQVPMRQSAEMDLASIAVAYTYLITLITAFPLIVIAANIPFYL 60

35 Query: 61 HINVDLLSLMQLPQHYEPASRLVADAPSKPSTGLIGFASLTAFWMSKSLTSIQKA 120  
 +I+++DLL LN++MLPK+I+ PAS + + PSKPS +IG A+LT WMS+SITSILQKA  
 40 Sbjct: 61 NIDIADLLRMKQNLPKDIFRPPASAIVENIPSKPSGSLGVATLTGLWIMSRSLTSIQKA 120

Query: 121 INKAYGVDCHRDVFVISRLVGVQTLTILPLTFLPISTPSKPVLIIVVMYDGLQTLTA 180  
 INKAYG CHRD I LVG+ T LILFLL P LIPS PSK +Q++ Y L D +T  
 45 Sbjct: 121 INKAYGASQHRDFFIGHLVGLTSLTILPLAFALIPISPSKAAIQVLDKHYHLSDNITT 180

Query: 181 WLNLAQPVFTLITFLGIGILYFLPNARIKRVYVIRPOTLFTSTFVIGFPSNLISQVLA 240  
 L L QP+T L IP+G+ +LYP+LN -I+K+RY++POTLP++FV+ F SML+ YV+  
 50 Sbjct: 181 IFILLIQPTVLIIFVGMILLYFLPNVKKIKIRYILPOTLFTSFVMTLHSLVGRYVY 240

Query: 241 RVEKMDIKTPGSVVIPIMLWFIPLAHIMLGAINASVCEIATGKIESRGGIMSLIQ 300  
 VE+VMVDIK FGSV+IFI+MILWFIPLA I+ILGAI NA+ QE++ GK+R R GD+++++  
 55 Sbjct: 241 NVERMDIKMFGSVMIFIMLWFIPLARILLGAI FNATYQMSLQKEGRSGMDIALLK 300

Query: 301 KS 302  
 K+  
 Sbjct: 301 KT 302

- 55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1428

A DNA sequence (GBSx1514) was identified in *S. agalactiae* <SEQ ID 4385> which encodes the amino acid sequence <SEQ ID 4386>. Analysis of this protein sequence reveals the following:

- 60 Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

-1576-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4200 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 **Example 1429**

A DNA sequence (GBSx1515) was identified in *S.agalactiae* <SEQ ID 4387> which encodes the amino acid sequence <SEQ ID 4388>. This protein is predicted to be methionine aminopeptidase (map). Analysis of this protein sequence reveals the following:

Possible site: 14

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2342 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9761> which encodes amino acid sequence <SEQ ID 9762> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25

>GP:AAC35914 GB:AF071085 methionine aminopeptidase A [Enterococcus  
 faecalis]  
 Identities = 101/207 (48%), Positives = 128/207 (61%), Gaps = 31/207 (14%)

30

Query: 1 NITLKSAREIEAMDRAQDFLASIHIGLRDIKPGVDMVEVEYVRRCKEENVLPLQIGV 60  
 NITLKS REIE MD +G+ LA +H LR IKRG+ W++E +VR + + QIG  
 Sbjct: 1 NITLKS PREIEMMOESGELLADVHRHLRTFFIKPGITSWDIEVFVRDFISHGGVAAQIGY 60

35

Query: 61 DGAVMDDPYATCCGLANDEVAAHAFPHRYHTLKQZLLKVDMLVSEPLDKSIVDVSLNFINV 120  
 +G Y YATCC +NDE+ H FFR LK GDL+KVID+ +  
 Sbjct: 61 EG----YKATCCSINDSI CHGFFRKVKVKGGLIKVDKVCVD----- 98

40

Query: 121 AQMKKYTETYSGLADSCWAVYAGVEVSQEVKDLMSVTEHMYTIGIKAVIGNRIGDIGA 180  
 G ++DSQW+Y VGE + E+ LM VT++A+Y+GIE+A +GNRIGDIG A  
 Sbjct: 99 -----LKGASISDSQWSYVVGESTPEIDRLMEVTKKALNLGIRCAQVGNRIGDIGA 149

Query: 181 IQVYASRRGYGVVRDLVGHGVGPNNHR 207  
 IQ Y E CNGVVRD VGHG+GPT+HE  
 Sbjct: 150 IQTVVSGEYGVVRDVPVGHGIGPTIHE 176

45

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4389> which encodes the amino acid sequence <SEQ ID 4390>. Analysis of this protein sequence reveals the following:

Possible site: 42

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2082 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

-1577-

Identities = 256/286 (89%), Positives = 273/286 (94%)

Query: 1 MITLKSAREIEMDRAGDFLAGIHIGLRDIIKPGVDMWVEVEYVRRCKENVLPLQIGV 60  
 5 Sbjct: 1 MITLKSAREIEMDRAGDFLAGIHIGLRDIIKPGVDMWVEVEYVRRCKENVLPLQIGV 60  
 MITLKSAREIEMDRAGDFLAGIHIGLRDIIKPGVDMWVEVEYVRRCKENVLPLQIGV 60

Query: 61 DGAAMDYPYATCCGLANDEVAHAFPHRYTLKQGDLLKVDMLSEPLDKSIVDVSLNFDNV 120  
 10 Sbjct: 61 DG +MDYPYATCCGLANDEVAHAFPHRY LK+GDLLKVDMLSEPLDKSIVDV++L+FDNV 120  
 DGHAMDYPYATCCGLANDEVAHAFPHRYLLKQGDLLKVDMLSEPLDKSIVDVAAALDFNV 120

Query: 121 AQMKKYTETYSGLLADSCWAYANGVSEVSKDILMSVTREMYIGIEKAVIGNRIGDIGAA 180  
 15 Sbjct: 121 PMKKYGTGYSGLLADSCWAYAVGTFSDEIKQLMDVTREMYIGIEKAVIGNRIGDIGAA 180  
 PMKKYGTGYSGLLADSCWAYAVGTFSDEIKQLMDVTREMYIGIEKAVIGNRIGDIGAA 180

Query: 181 IQDYAESRGYGVVRLDVGHGVGPTMHEEPMPVNYGTAGRLRLKGMVLITIEPMINTGTW 240  
 20 Sbjct: 181 +Q+YAES GYGVVRLDVGHGVGPTMHEEPMPVNYGTAGRLRLKGMVLIT+EPMINTGTW 240  
 VQDYAESFGYGVVRLDVGHGVGPTMHEEPMPVNYGTAGRLRLKGMVLITVPEMINTGTW 240

Query: 241 EIITDMKTGWAHKTLDGGLSCQYSHQFVITKDGPVILTSQGEERTY 286  
 25 Sbjct: 241 EIITDMKTGWAHKTLDGGLSCQYSHQFVITKDGPVILTSQGEERTY 286  
 EIITDMKTGWAHKTLDGGLSCQYSHQFVITKDGPVILTSQGEERTY 286

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 1430

A DNA sequence (GBSx1516) was identified in *S.agalactiae* <SEQ ID 4391> which encodes the amino acid sequence <SEQ ID 4392>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3473 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9759> which encodes amino acid sequence <SEQ ID 9760> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BA06894 GB:AP001518 unknown conserved protein [Bacillus halodurans]  
 Identities = 158/431 (36%), Positives = 270/431 (61%), Gaps = 6/431 (1%)

Query: 6 SKHQILEYLENLAVGRVSRISNHLKVSQDTAYRAIKHAENRGIVETRRSGSTVRVA 65  
 45 Sbjct: 3 +KH+IL+Y+ NL VG++GVR I+ L+V+S+GTAYRAIKHAEN+G+V T R GT+R+ 62  
 TKGHEQLQYITNLEVGEEKISVRRIRAKDLQVSGSTAYRAIKHAENQGLVSTIERVGTIRIE 62

Query: 66 QKAKVNIKLTAYAKIRISDSQVAGIEGLSKESPKFSIGAWTHRNIEKYLVOGELLIVG 125  
 50 Sbjct: 63 +K K NIEKLTAY+ I D QV+ G +GL K ++F IGM + +Y+ G LLIVG 122  
 KKKQENIEKLTAYAKVNVIVDQVLAGRDGLKHTLNRFVIGAMKLDMMRYVEPGNLLIVG 122

Query: 126 DRDEIQLALQHQNAIILVTGGFNPSPVCKADKLQIPVNVITHTYTFVSTMDIELNSA 185  
 55 Sbjct: 123 +R ++ +AL+ A+L+TGGF+ S +LAD+L +PV+ T YDTFTV+TMIN ++ 182  
 NRYQVHQIQAEGAAVNLITGGFTSDRAIKLADLDELPLVISTSTYDTFTVATMINRAIYDQ 182

Query: 186 KIRIDLKTVQEVYQSQMDYGLAQDITVKEPNLVLKQTNVRFPIVQNVVGVGVSD 245  
 60 Sbjct: 183 I+ ++ V+ + D ++ ++ V +++ L ++T + R+P+++ + G+V+ +D 242  
 LLIKIEITLVDDILIPQDTYMTTENVVGVKHELNKETHSRYPVIDENMKIQGWAAKD 242

Query: 246 ILGHKEVKLATVMSKNIIVAKPMISLANISQKMFEDLAWQPVVSDPELIGVITTRQA 305  
 65 Sbjct: 243 VLNASRHTPIEKVMTKNPTVSEKTSVAAVAHVWVGIRLLFPVIDSHKRLIGVVSQDV 302  
 +L + + VM+KN I R S+A ++ M+R+ ++PV+ +L+GV+R+ 302

-1578-

Query: 306 VENVLSMSQ----GTLLTYTSDQILSNLQIEDG-HFSLVLRPAMIDHTGSLTQGVLTPEFL 359  
 ++ L M Q G + L+ + G + + P M + G++ GV+T +  
 5 Sbjct: 303 LKALQMIQRPHVGTETIEDLWINGLNKSSSQGSDSYVEVETPQMTNQLGTTISGVMVSLV 362

Query: 360 KRICIRVLTRKHQRISIVVKQNTLYFLQPVQIDETIMVPTTISEKKRRATDLLEKLKLNK 419  
 E RVL + + +VV+ +ILYFL+PVQID + + P ++ R+ +D+E+ E +  
 10 Sbjct: 363 IBSGSRVLRKYYKKGDIIVENTILYFLQPVQIDSLRTIRPVLEIGRKHGKIDVEMYHGE 422

Query: 420 IIAKAMIAVKI 430  
 I+ KA+ +I  
 10 Sbjct: 423 IVGKALPMAQI 433

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4393> which encodes the amino acid  
 15 sequence <SEQ ID 4394>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 20 bacterial cytoplasm --- Certainty=0.3011 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 267/431 (61%), Positives = 351/431 (80%)

Query: 1 MIIIVMSKHQRILEYLENLAVGKRVSVRSISNHLKVSOGTAYRAIKEAENRGIVETKPRSG 60  
 ++IIMKHQ+IL+YLE LA-QK+VSVRSISNHLKVSODTAYRAIKEAENRGIVET+KPRSG  
 30 Sbjct: 1 VIIIMSKHQIDILYLEKLAIGKKVSVRSISNHLKVSOGTAYRAIKEAENRGIVETKPRSG 60

Query: 61 TVRVAQKAKVMISKLTYSIARISSDQVVGIEGLSKFEPSKPSIGAMTHRTIKRYLVQGG 120  
 TVR+ +K +V I++ITY+IARISSD+V+AG GL EFS+PSIGAMT +RI +YLV+GG  
 30 Sbjct: 61 TVRIEKKGRVRIIDRLTYSSIARISSDSEVLGAGLGHFEPSPSIGAMTQQNIRRYLVKGG 120

Query: 121 LLTVGDRDEIOHLALCHQNALIVTGQPNVSPSVCRADKLQIPMVTHYDTFTVSTMINH 180  
 LLTVGDR+ IQ LAL++ NAILAVTGGF VS V +A+ +IPMVTHYDTFTV+TMINH  
 35 Sbjct: 121 LLTVGDRRETIQALLALENNAILVTGGFPVSKRVISMANQRIPMVTHYDTFTVATMINH 180

Query: 181 TISNAKIRTDLKTVEQVYQSQMDYGLAQDDTVKEFNHLVKQTKNVRFPVQANVVG 240  
 LSN +I+TDLKTVEQV DYGL+D +V+EPN L+K+T+ VRFP++ V+GV  
 40 Sbjct: 181 ALSNIRIKTDLKTVEQVMIPITIDYGLCEDSSVEEPNLTLLIKTRQVRFVLDYKRVIGV 240

Query: 241 VSVQDILGKDEKVKLATVMSKNIIVAKFRMSLANISQKMFEDLNMPVSDDFELLGVI 300  
 VS++D++ + KL VMSKN I A+P SLANISQKMFEDLNMPV++ LLG+I  
 45 Sbjct: 241 VSMRDVVDQLFTTKLTKVMSKNPITARNTSLANISQKMFEDLNMLPVDRENNLLGMI 300

Query: 301 TRRQAVENLSMSGTDLTYTSDQILENLQIEDGHFSLVLRPAMIDHTGSLTQGVLTPEFL 360  
 TRRQA+ENL Q + YTS+QILENL+ ++ +VEP MID G++ GV+REFLK  
 50 Sbjct: 301 TRRQAVENLNHQPNNPPTYSQILENLBETVDYQVVEPTMIDSAQMSNGVISEFLK 360

Query: 361 EICIRVLTRKHQRISIVVKQNTLYFLQPVQIDETIMVPTTISEKKRRATDLLEKLKLNK 420  
 EI IR LT+KHQ++I++QM +YFL +QI++ + P II+E RR +T+D+E+ +++++  
 50 Sbjct: 361 EISIRALTRKHQKNIIIEQMNVYFLHAIQIEDELKIYPKIIITENRSSTIDIEIFVDDQV 420

Query: 421 IAKAMIAVKIN 431  
 IAKA+I KIN  
 55 Sbjct: 421 IAKAIITTKIN 431

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

-1579-

**Example 1431**

A DNA sequence (GBSx1517) was identified in *S.agalactiae* <SEQ ID 4395> which encodes the amino acid sequence <SEQ ID 4396>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2837 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04556 GB:AP001510 unknown conserved protein [Bacillus halodurans]
Identities = 56/185 (30%), Positives = 86/185 (46%), Gaps = 4/185 (2%)

Query: 7 MDIWNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRINLQFI-PPVQTSKKGSDFLV 65
M+I G +ETE+LR D A+ AS +++ + S K+8+ L
Sbjct: 1 MEIEDIYGLDPLFLETERLRRLKPKFKDDAAIYDYASNEQVTKYVLWETHQSIRKSEAPLA 60

Query: 66 HSPMK---EPLGVWALEDKVSHKMGFVIRFENIDLSKKTAEIGYFLKSSWGQIMTECL 122
+ K ++ WAIR K + +MG + P KTAE+GY L E WQGIMTE +
Sbjct: 61 FALNKYDEKLVSPWAIELKRNERMIGTVDPVWKKPKDKTAEALGYVLESDYWGQIMTEAV 120

Query: 123 KTLSPFAPREFGMKLLIVTHKENIASQKVALKAHFKQSRSPKGSDDRYTRIRIDYIEFOL 182
L F F +++++ ENI+S +V KA + + + RD+ + +
Sbjct: 121 NALVPEFGNNMELERIQAKCFENISSARVMKAGLIYBOTHRRRAIYVGKHAHRDFKVAI 180

Query: 183 TRGDY 187
R DY
Sbjct: 181 IREDY 185

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 667> which encodes the amino acid sequence <SEQ ID 668>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1096 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 94/177 (53%), Positives = 117/177 (65%)

Query: 7 MDIWNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRINLQFI-PPVQTSKKGSDFLV 66
MDIWT L FAF ET V LRP Y D F+ + + NL ++PF Q +K SD+LLVH
Sbjct: 1 MDIWTKLAVFAFFETPKVILRPFYEDHMDYFVNVDTKNLYYVFPBQKTAASDYLVLH 60

Query: 67 SPKKEPLGVWALEDKVSHKMGFVIRFENIDLSKKTAEIGYFLKSSWGQIMTECLTSL 126
SF+K PLG WALEDK +H+ G IR E+ D + A+IGYFL + WQGIMTE + L
Sbjct: 61 SFIKPLGVWALEDKATHQVIGSIRIERHYDAKTRADIGVFLNYAFWQGIMTEVVIKL 120

Query: 127 FFAFPREFGMKLLIVTHKENIASQKVALKAHFKQSRSPKGSDDRYTRIRIDYIEFOLT 183
+ + F EPG+ L I+TH EN ASQKVA KA F+ FKGSDDRYTRIRIDYIEFOLT 177
Sbjct: 121 YLSPHEPLKTLRIITHLENKASQKVAKKAGFQLKTCFKGSDDRYTRIRIDYIEFOLT 177

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1580-

## Example 1432

A DNA sequence (GBSx1518) was identified in *S.agalactiae* <SEQ ID 4397> which encodes the amino acid sequence <SEQ ID 4398>. This protein is predicted to be UDP-N-acetylglucosamine-1-carboxyvinyl transferase (murA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 61
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -5.63    Transmembrane    25 - 41 ( 24 - 42)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAF96297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
      transferase [Listeria monocytogenes]
      Identities = 240/412 (58%), Positives = 303/412 (73%), Gaps = 2/412 (0%)

20  Query: 3 KIIINGGKQLTGEVAVSGAKNSVVALIPATILADVVLDGVPASIDVDSLVDMETGA 62
      K+II GKK+L G + V GAKNS VALIPA TIa+ VVL+G+P ISDV +L +I+E +G
      Sbjct: 20 KLIRINGGKLAGITLQVDGAKNSAVALIPAALIAESVVLGLEDISDVHTLNLIELG 79

      Query: 63 KIKRYGETLEIDPOGVKDIPMPYGKINSLRASYTFGSELLGRYQATLGLFGGCDLGRFP 122
      ++ +T IDP + +P+P G + LRSASY G++LGR+ +A +GLFGGC LGPRP
      Sbjct: 80 TVRYDNKTAVIDPDTMISMPLESGNVKLRASYTYLGMALGRFKKAVIGLFGGCTLGFRP 139

      Query: 123 IDLHLJAFEMGASVSYEGDSMRLATNGKPLQGANIYMDTVSVGATINTIIAAKANGRT 182
      ID H+K FEA+GA V+ E ++ L + L+GA IY+D VSVGATIN ++AA +A G+T
      Sbjct: 140 IDQHKIGFEALGAKVTNEQGAITYLRAD--ELKGARIYLDVSVGATINIMLANVRAKGT 197

30  Query: 183 VIEAAREPEIIVATLLNNMGAIHIRGAGTDVITIEGVKSLHGTNRHQVPRIRIAGTYIA 242
      VIEAAR+PEIIVATLL NMGA I+GAGTD I I GV+ LHG H +IPRIRIAGT++
      Sbjct: 198 VIEAAREPEIIVATLLNMGAIIKAGAGTDVTRITRGVHGLGCHHTIIPRIRIAGTFMV 257

35  Query: 243 MAJAIIGRIKIVNVLVYHLESTIAKLDEMGVMTVEEDSIFVEEQERLKAIVSIRTSPTPG 302
      +AA G+G+++ NV+ HLE IAKL EMGV M +EED+IFV E E++K V IRT YPG
      Sbjct: 258 LAASGKGVRLENVPIPHLEGITIAKLDEMGVPMIDIEDAIFVGEVEKIKKVDIKTYAYG 317

40  Query: 303 FATDLQQPLTPLLITABINGSLDITTYEKRNVHVEPLARMGANISTLGGKIVYSGPNQLS 362
      F TDLQQPLT LL ABG+ +DTTY R H+ E+ RMG G V +GP QL
      Sbjct: 318 FPTDLQQPLTALLTRABGSSVITOTIYPSRFCHARTIRMGKFK+EGRSVAVINGPVLQ 377

      Query: 363 GAPVKATDLRAGALVIAAGLMAEGRTEITNIEFILRGYSNIEKLITSLGADI 414
      G+ V ATDLRAGALVIA L+A+G TEI +E I RGSY IIEKL++GA+I
      Sbjct: 378 GSKVTATDLRAGALVIAALLADGETEIHGVHIERGRYSKITEKLSAIGANT 429

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4399> which encodes the amino acid sequence <SEQ ID 4400>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 21
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.70    Transmembrane    25 - 41 ( 23 - 45)

55  ----- Final Results -----
      bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

60  >GP:AAF96297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
      transferase [Listeria monocytogenes]
      Identities = 244/412 (59%), Positives = 302/412 (73%), Gaps = 2/412 (0%)

```

-1581-

Query: 3 KIIINGGKALSGEVAVSAGKNSVVALIPAILLADDTIVLDGVPAISDVDSLIEIMELMG 62

5 Sbjct: 20 K+II GOK L+G + V GAKNS VALIPA ILA+ V+L+G+P ISDV +L I+E +G  
KLIIRGGKKLAGTIVDQVAGKNSVVALIPAILLADDTIVLDGVPAISDVDSLIEIMELMG 79

Query: 63 TVNYHGDLEIDPRGVQDIIMPYGGKINSLRASYYPYGSLLGRFGQAVGLPGGCDLGRP 122  
TV Y T IDP + +P+P G + LRASY G++LGRF +AV+GLPGGCD LGHRP  
Sbjct: 80 TVRYDNKTAVIDPTDMISMFPLPGSNVKKLSRASYIMGAMLRGFKKAVIGLPGGCDLGRP 139

10 Query: 123 IDLHLKAFAMGVSVEYSGNNMLSTNGQKIHGAHYMDTVSVGATINTMVAATKAQSKT 162  
ID H+K FEA+G +V+ E + L + ++ GA IY+D VSVGATIN M+AA +A+GKT  
Sbjct: 140 IDQRIKGFEALGAKVTNSGAIYLRAD--LKGARIYLDVVSVGATINIMLAIVRAKGT 197

15 Query: 183 VIENAAEPEIIDVATLLNNMGAHIRGAGTDITTIQGVKLGHTRHQVIFDRIEAGTYIA 242  
VIENAA+EPEIIDVATLL NMGA I+GAGTD I I GV+ LHG H +IPDRIEAGT++  
Sbjct: 198 VIENAAEPEIIDVATLLNNMGAHIRGAGTDITIRITGVHRLGCHHTIIIPDRIEAGT+V 257

20 Query: 243 LAAAIGKGVKITNVLYHEHLSFIAKLEMGVMTVEEDAIFVEKQBSLKAITIKTSYPG 302  
LAAA GKGVI+I NV+ HLE IAKL EMGV M +EEDAIFV + E +K + IKT YPG  
Sbjct: 258 LAAASGKGVRIENVIPFLEGI IAKL EMGV PMDEDAIFVGEKIKVDIKTYAPG 317

Query: 303 FATDLQQPLTPLLLKADGRGTTIIDTYEKRINRIVPELMRMGADISVIGGQIVYQGSRLT 362  
F TDLQQPLT LL +A+G I DTII R H+ E+ RMG + G V GP +L  
25 Sbjct: 318 FPTDLQQPLTALLTRAEGSVITDTIYPSRFGIHAETIERMGKFKLEGRSAVINGQQLQ 377

Query: 363 GAQVKATDLRAGAALVTAGLIAEGKTEITNIEPILRGYASIIAKLTALGADI 414  
G++V ATDLRAGAALV L+A+G+TEI +K I RGT+ II KL+A+GA+I  
30 Sbjct: 378 GSKVTATDLRAGAALVIAALLADGETIEHGVHEIRGYSTIEKLGAIGNI 429

An alignment of the GAS and GBS proteins is shown below.

Identities = 344/419 (82%), Positives = 394/419 (93%)

35 Query: 1 MRKIIINGGKQUTGEVAVSAGKNSVVALIPAILLADDTIVLDGVPAISDVDSLIDMETM 60  
Sbjct: 1 MRKIIINGGK L+GEVAVSAGKNSVVALIPA ILADD+V+LDGVPAISDVDSL++IME M  
MRKIIINGGKALSGEVAVSAGKNSVVALIPAILLADDTIVLDGVPAISDVDSLIEIMELM 60

Query: 61 GAKIKRYGETLEIDPGVKDIMPYGGKINSLRASYYPYGSLLGRFGQAVGLPGGCDLGP 120  
GA + +G+TLEIDP GV+DIMPYGGKINSLRASYYPYGSLLGR+GQA +GLPGGCDLGP  
40 Sbjct: 61 GATVNYHGDLEIDPRGVQDIIMPYGGKINSLRASYYPYGSLLGRFGQAVGLPGGCDLGP 120

Query: 121 RPIDLHLKAFAMGVSVEYSGNNMLSTNGQKIHGAHYMDTVSVGATINTIIAAAKANG 180  
RPIDLHLKAFAMG VSYEG+M L+TMG+ + GA+IYMDTVSVGATINT++AA KA G  
45 Sbjct: 121 RPIDLHLKAFAMGVSVEYSGNNMLSTNGQKIHGAHYMDTVSVGATINTMVAATKAQ 180

Query: 181 RTVIENAAEPEIIDVATLLNNMGAHIRGAGTDVITIEGKSLHGTTRHQVIFDRIEAGTY 240  
+TVIENAAEPEIIDVATLLNNMGAHIRGAGTD+ITI+GV+ LHGTRHQVIFDRIEAGTY  
Sbjct: 181 RTVIENAAEPEIIDVATLLNNMGAHIRGAGTDITTIQGVKLGHTRHQVIFDRIEAGTY 240

50 Query: 241 IAAAAGIKGVKITNVLYHEHLSFIAKLEMGVMTVEEDAIFVEKQBSLKAITIKTSYP 300  
IA+AAAIG+K+TNVLYHEHLSFIAKLEMGVMTVEEDAIFVE+QE LKA++IKTSYP  
Sbjct: 241 IALAAAIGKGVKITNVLYHEHLSFIAKLEMGVMTVEEDAIFVEKQBSLKAITIKTSYP 300

55 Query: 301 PGFATDLQQPLTPLLLTABGNSLDTIYTKRINRIVPELMRMGADISVIGGQIVYQGSRL 360  
PGFATDLQQPLTPILL A+G G+++DITYEKR+NRIVPEL RMGA+IS +GG+IVY GL++  
Sbjct: 301 PGFATDLQQPLTPLLLKADGRGTTIIDTYEKRINRIVPELMRMGADISVIGGQIVYQGSRL 360

Query: 361 LSGAPVKATDLRAGAALVTAGLIAEGKTEITNIEPILRGYASIIAKLTALGADTQLIED 419  
L+GA VKATDLRAGAALV AGL+REG+TEITNIEPILRGY+II KIT+LQADTQL+H+  
60 Sbjct: 361 LTGAQVKATDLRAGAALVTAGLIAEGKTEITNIEPILRGYASIIAKLTALGADTQLIED 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1582-

**Example 1433**

A DNA sequence (GBSx1519) was identified in *S. agalactiae* <SEQ ID 4401> which encodes the amino acid sequence <SEQ ID 4402>. This protein is predicted to be thiamine phosphate pyrophosphorylase (thiE). Analysis of this protein sequence reveals the following:

5       Possible site: 55  
       >>> Seems to have no N-terminal signal sequence

10       ----- Final Results -----  
               bacterial cytoplasm --- Certainty=0.0422 (Affirmative) < succ>  
               bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
               bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15       >GP:AAF25544 GB:AF109218 ThiE [Staphylococcus carnosus]  
       Identities = 98/200 (49%), Positives = 140/200 (70%), Gaps = 1/200 (0%)

      Query: 5   LKLNYFVCGITVDCSR-RNILT'VVEALQAGITLQPREKGFALQOKEKIANAKQLQILCK 63  
               L +YF+CGT D + I V++EAL+ GITL+QFRKKG A G++K++AK+LQ LCK  
       Sbjct: 7   LNVYFIQGTQDIPGRTIQEVLKEALEGGITLQPREKGNKATQDQKVALAKELQALCK 66

20       Query: 64   QYQVFFIIDDIDLVLIDADGLHIGQNDLPVDEARRRLPKIIGLSVSTMDERYQKSQLS 123  
               Y VPFI++DD+ L S IDADG+H+GQ+D VD+ R KIIGLS+ ++E S L+  
       Sbjct: 67   SYIVPFIIVDDVALAEIADADGIHVQDDDAVDDFNNRFEKGIIGLSIGNLEELASDLT 126

25       Query: 124 VVDYIGIGFPENPTQSKADAKPAVGNRTTKAVREINQDIPIVAIGGITSDFVHDIESGAD 183  
               VDYIG+GP T SK DA VG + + +R+ D+PIVAIGGI+ D V ++ ++ AD  
       Sbjct: 127 VVDYIGVGPIFATPSKDDASEFVGPKMIETLRKEVGDLPIVAIGGISLNDVQEVAKTSAD 186

30       Query: 184 GIAIVISAISKANHIVDATRQ 203  
               G++VISAI++ H+ + +  
       Sbjct: 187 GVSVISALARSPhVTETVHK 206

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 35       vaccines or diagnostics.

**Example 1434**

A DNA sequence (GBSx1520) was identified in *S. agalactiae* <SEQ ID 4403> which encodes the amino acid sequence <SEQ ID 4404>. This protein is predicted to be hydroxyethylthiazole kinase (b2104). Analysis of this protein sequence reveals the following:

40       Possible site: 54  
       >>> Seems to have a cleavable N-term signal seq.  
               INTEGRAL   Likelihood = -4.94   Transmembrane 198 - 214 { 194 - 217}

45       ----- Final Results -----  
               bacterial membrane --- Certainty=0.2975 (Affirmative) < succ>  
               bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
               bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8805> which encodes amino acid sequence <SEQ ID 8806>  
 50       was also identified. Analysis of this protein sequence reveals the following:

      Lipop: Possible site: -1   Crend: 7  
       McG: Discrim Score: -2.93  
       GvH: Signal Score (-7.5): 1.61  
               Possible site: 39  
 55       >>> Seems to have no N-terminal signal sequence



-1583-

ALOM program count: 1 value: -4.94 threshold: 0.0  
 INTEGRAL Likelihood = -4.94 Transmembrane 183 - 199 ( 179 - 202)  
 PERIPHERAL Likelihood = 2.49 151  
 modified ALOM score: 1.49

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.2975(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25543 GB:AF109218 ThiM [Staphylococcus carnosus]  
 Identities = 114/253 (45%), Positives = 160/253 (63%), Gaps = 1/253 (0%)

Query: 18 LEQLKEVNPLTICITNNVKNFTNGLLALGASPMSECTEDLEDLLKVDALLINIGTL 77  
 L+Q++ +PL IC TN+VKNFTNGLL+LGASP MSE ++ ED VA ++L+INIGTL  
 Sbjct: 5 LQQRTEHPLVICYTNVKNFTNGLLALGASPTMSERAPQEAEDFPYVAGSVLINIGTL 64

Query: 78 TKESWQLYQEAIKIANKNQVPPVLDFAAGASRPLEVSLDLLKNYSISLITNGSGEIAA 137  
 TK E KIAN+ + P+V DPVA GAS++R + LK +++ GN SEI A  
 Sbjct: 65 TKHHEHAMLENKIANETETPLVDFPVAVGASKYKDPCKYPLKKIKPTVIKNGASEILA 124

Query: 138 LIGKQASKGADGGKVADLESIAVKANQVDFPVPVVVTGETDAIAVRGEVRLQLNGSPLMP 197  
 LI + KG D+ IA KA + + ++TGTD I +V L NGS +  
 Sbjct: 125 LIDDTATMKGTDSADNLDVVDIAEKAYEYQTAILITGETDVIQDNKVVILNKGSHFLA 184

Query: 198 LVITGTCLLGAVLAAPFGSSDRSDDLACLTEAMTVYNVAGEIAKVGAKGVGSPQVAPL 257  
 +TG GCLLGAV+ AP+ + + L EA++VYN+A E AE+++ KG G+P F+  
 Sbjct: 185 KITGAGCLLGAVGAPL-FRNTHPSETLLEAVSVYNIAERABQLSDSGPGTFLQTFI 243

Query: 258 DALSQMKSEIMD 270  
 DAL ++ S+ ++  
 Sbjct: 244 DALYRIDSDAVAE 256

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8806 (GBS398) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 6; MW 31.8kDa).

The GBS398-His fusion product was purified (Figure 214, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 314), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1435

A DNA sequence (GBSx1521) was identified in *S.agalactiae* <SEQ ID 4405> which encodes the amino acid sequence <SEQ ID 4406>. This protein is predicted to be ThiD (thiD). Analysis of this protein sequence reveals the following:

Possible site: 44  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1584-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25542 GB:AF109218 Thid [Staphylococcus carnosus]
Identities = 139/258 (53%), Positives = 186/258 (71%), Gaps = 4/258 (1%)

5 Query: 8 LTIAGTDPGSGAGINADLKTFQARRTYGMVVTSVVAQMTQVSRGVQHIETAIIDQQLAC 67
      LTIAGTDP+GGAG+MADLK+F A YGMA +TS+VAQMT GV+ + +++ + +GL
Sbjct: 8 LTIAGTDP+TGGAGVMADLKSFIACGVYGMALITSIVAQMTKGVQHIHNLIDITVWLBQLDS 67

10 Query: 68 VYDDIKPKAVKTCMLAERETISLVASYLKKYPC-PYVLDPMVAISGHRLLIDDAVEALK 126
      ++DD P+A+KIGM+A +S + L+ SYL+KYP PIV+DFVM+A SG L+D AL+
Sbjct: 68 IFDDELPCQAIKTMGLATKEMMELISYLEKYPIPIVVIDPVMALAKSGDSLMDAGKHALQ 127

15 Query: 127 EDLLPLATIITPNLPEASVLGVYDLSDEVSIIKAGYDIQKYSVRNVLIKGSHLD--GLA 184
      E LLPLA + TPNLPEAS +VG+ L E +I KAG + + +V+IKGSH+ +A
Sbjct: 128 RILLPLADVATPNLPEASEIVGFKLUTRAIKKAGDIFINETSIGSGGVVINGKHIEDKNIA 187

20 Query: 185 KDYLFLRKEGLITLTSNQRINTHTHTGCTFPAVVAELAKGQSLIMAVSTANSFITS 244
      KDYLF K+GL ++R +I HTHGTGCTF+AV+ AELAKG++I AV AK FI +I
Sbjct: 188 KDYLF-TRDGLKVFESERYDTKHTHTGCTFSAVITAEALAKGKTIYBAVKKARDFTALSI 246

25 Query: 245 ETAPELGLGNGPVNHTSY 262
      + FE+G G GPVNH +Y
Sbjct: 247 KTFPEIGQGRGPVNHFAY 264
```

25 There is also homology to SEQ ID 4408.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1436

30 A DNA sequence (GBSx1522) was identified in *Sagalactiae* <SEQ ID 4409> which encodes the amino acid sequence <SEQ ID 4410>. This protein is predicted to be TenA (tenA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2242 (Affixmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25541 GB:AF109218 TenA [Staphylococcus carnosus]
Identities = 78/213 (36%), Positives = 127/213 (59%), Gaps = 6/213 (2%)

45 Query: 14 IQSIYODPPTIQGIKGRLDHVDICHYLQADNIYLGKFPADIVYALCLAKSDNIIRDKQFPLRQ 73
      I IYOD FIQ ++KG+ + + YL+AD YL +FA+IYAL + L+ +F +F+Q
Sbjct: 15 IDRSIQDHFIIQELKGDUIKKEALRQYLADASYLERFANITYALLIPMDLESVRFLVDQ 74

Query: 74 IDFTLNRELAIDGEGPHQALAAVINRSYQDILKGVVNPSPADHYIKHMYFHYF-ENKLAGA 132
      I F +N E+ H+ +A Y +Y +I++K VN PS DHYIKHMY++ Y A A
50 Sbjct: 75 IOFIVNGEVE----AHEYMADYIGSNYNRIVQKKVNPSPGDHYIKHMYNYVAHENAAYA 130

Query: 133 LAANSFPCWIIYHQLAKKIIRKNOFLNGPNFNNWITFYANDTVEELMENYFRM+DYAYQNL 192
      +AAM+PCP+Y +AK+ +++ + W FY N ++L+E +M+ N+
55 Sbjct: 231 LAAMAPCPYVYAMIAKRAMKDPLNKSILAKWFEFY-NTEMDFLIEVLDLQNLQNTAM 189

Query: 193 SKEKQADLVDAFVKSCHERRFPQMAINCENWE 225
      S+ ++ ++ + +++S HE FF MA EKW+
Sbjct: 190 SRTEKNEVRENILQSTVHELNFNFMAYTSKQW 222
```

-1585-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1437

- 5 A DNA sequence (GBSx1523) was identified in *S.agalactiae* <SEQ ID 4411> which encodes the amino acid sequence <SEQ ID 4412>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
10  INTEGRAL Likelihood = -7.06 Transmembrane 43 - 59 ( 36 - 63)
    INTEGRAL Likelihood = -2.55 Transmembrane 92 - 108 ( 92 - 112)
    INTEGRAL Likelihood = -1.49 Transmembrane 135 - 151 ( 135 - 151)
    INTEGRAL Likelihood = -1.06 Transmembrane 69 - 85 ( 69 - 85)
    INTEGRAL Likelihood = -0.22 Transmembrane 216 - 232 ( 216 - 232)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20  The protein has homology with the following sequences in the GENPEPT database.
    >GP:CAA91230 GB:Z56283 orf2 [Lactobacillus helveticus]
    Identities = 46/215 (21%), Positives = 96/215 (44%), Gaps = 3/215 (1%)

25  Query: 21 AITFLCILLIPTFSFPTLRLETSLFLIIVVTLQCFVSKLTKWAKNLIISFVMSLSFL 80
    ++ F+ I + S L T+L+ + + ++ +K + + F+ ++F
    Sbjct: 4 SLKFLAFLIISLEISLKASULTNLIIVIAFALIYLVTRIKINELILLIAVFIASPTIFA 63

    Query: 81 GTYFWGKLPHQFVLASLVACRPLIFMNVGLLPHASHSNYDFIESLKQTPKVPBHFAYQIF 140
    +++ F + +L+ R ++ + + DF SL Q +PS FRIG+
    30  Sbjct: 64 TLFWFSPTDAYYANWL-STRVYVYTLTIACVTRNTATDFARSLQNHLHPSKFAYGVL 122

    Query: 141 AVFILLLLKIKLYQRMRLAPRLKNQVTWALSRLILSVLLKTIYVWQLELANLSKQFEV 200
    A N++P +K ++ R + ++ 3P L +L + + L M S G+
    35  Sbjct: 123 AAINIIPRMKTAVKQIRTSAMNRGMYLSFWSFVLYKAILVALNSADNLAQGMESHGYVE 182

    Query: 201 QKERTHASTYTVRPRDYSL-LQMSILLIGM-IFK 233
    G++R P+ +D+ + + IL++I + IFK
    Sbjct: 183 GQKRATIVAIPLTKKDWLIFPFTLLIIVNISLPIFK 217
  
```

- 40 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8807> and protein <SEQ ID 8808> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 0
McG: Discrim Score: 4.50
GvH: Signal Score (-7.5): -0.2
    Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 5 value: -7.06 threshold: 0.0
50  INTEGRAL Likelihood = -7.06 Transmembrane 43 - 59 ( 36 - 63)
    INTEGRAL Likelihood = -2.55 Transmembrane 92 - 108 ( 92 - 112)
    INTEGRAL Likelihood = -1.49 Transmembrane 135 - 151 ( 135 - 151)
    INTEGRAL Likelihood = -1.06 Transmembrane 69 - 85 ( 69 - 85)
    INTEGRAL Likelihood = -0.22 Transmembrane 216 - 232 ( 216 - 232)
    PERIPHERAL Likelihood = 2.65 170
    modified ALOM score: 1.91
  
```

\*\*\* Reasoning Step: 3

-1586-

----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1438**

A DNA sequence (GBSx1524) was identified in *S.galactiae* <SEQ ID 4413> which encodes the amino acid sequence <SEQ ID 4414>. Analysis of this protein sequence reveals the following:

Possible site: 24

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3007(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15

The protein has homology with the following sequences in the GENPEPT database.

20

>GP:CAA91229 GB:Z56283 orf1 [Lactobacillus helveticus]  
 Identities = 123/424 (29%), Positives = 200/424 (47%), Gaps = 48/424 (11%)

25

Query: 17 LPDEVTFSINPGERILISGVSGCKSTLALLLSGL--KESGK--GQVLINGSLIEPSDVG 72  
 L +++ ++ PG +LI G +GGCKSTL ++GL K +GK G++ L+G  
 Sbjct: 12 LINQNMNIAPGPHILLI-GFTGGCKSTLLKTIAGLYPKYAGKLTGKIDLANGQ----KAA 65

30

Query: 73 FLFPQNDQLFCNDTVAHELYFILENLQIEPFGQDRSEFVLAQVGLKGPQRLIYTLSSQ 132  
 +FQN QF N T E F LENLQI+ + + + + ++ I TIS G  
 Sbjct: 66 MFPQNAABQFTMTTTPREIIFALENLQIKADYDLHIKKAVEFTKIADLDQKINTLSGG 125

35

Query: 133 EKQRALATIFLKSFKLIIIDBAFANLDOBSAQILLQVLVNIQANNSQMLVIDHILITY 192  
 ++Q +ALA + + +LDE FA+ D + L++ + + ++ +I+ DH++ Y  
 Sbjct: 126 QQQHVALAVLIAMQVDFVFLDEFFASCDFNTRHFLIEKLASIAETGRT-IILSDHVLDEY 184

40

Query: 193 QDIMCHYFWLEKRLTRVNFYMLNRLNVFLEKKSHN-----TGDKLLSTKOPVK- 243  
 + I DH + E + + N+L F+ K+ H TG + + Q+K  
 Sbjct: 185 EKICDHLYTFEGKTVKELSANENKIL--FKQKQFHEQSYSFALPTGTPVPELANKTQIKQ 242

45

Query: 244 ----LKNKPFISYLDLFLASGERICLDGPGSGVGKSSLPWGLLLNRTKQK-----KQ 291  
 L +NK Y G+ + G +GVGK+SLF + + KQ + +  
 Sbjct: 243 NRLLKQNKLTQY-----GKTTLTGSGNVGKTSLFKAMTYMIPYKGNFTYLDNEISK 295

50

Query: 292 FTERKQIP-ISPLFQNPFLDQFISTVYDEIPOVCKDGN-----KARDILETILMDKQK 344  
 +RK + I+ FQ DQF+ TV DEI KD N K + LE + L  
 Sbjct: 296 IKRKYLSQIAQFPQKASQDPLTVTVKDEIELSKDRNHPPTDAKIDEWLEKIQKQILD 355

55

Query: 345 FSPFLSGQQQRRLAIGSLIADSKILLIDEPYTGQDAYHANMITLLSYCHRNICQVI 404  
 + IS GQQ++L I + L + +LL+DEP G D +++ L+ K +  
 Sbjct: 356 QVVYSLSGQQQKQKILMLMTKHNVILIDEPLSGLDHESVDVLVQMQCBQKQLQIFL 415

60

Query: 405 FTSH 408  
 SH  
 Sbjct: 416 IISH 419  
 Identities = 44/185 (23%), Positives = 83/185 (44%), Gaps = 24/185 (12%)

Query: 28 GERILISGVSGCKSTLALLLSGLKESGKQVLINGSLIE-----SDVGFPLQNPDLQ 81  
 G+ LI+G +G GK++L ++ + L+ + + S + FQ Q  
 Sbjct: 256 GKTTLTITGSGNVGKTSLFKAMTYMIPYKGNFTYLDNEISKIKRKYLSQIAQFPQKASQ 315

Query: 82 PCMDTVAHELYFILENLQIEPFGQDRSEFV-----LAQVGLKGPQRLIYTLSSQ 133  
 F TV E+ +DR+ F L ++ LK ++++Y+LS G+

-1587-

Sbjct: 316 FLTIVTKDRIEL-----SKDRNNFFTDKIDENLEKILQKQHLQVQVYSLSGGQ 365

Query: 134 KQRALATIFLSPKGLIILDEAFANLDQESASQLQLVLVNYQANNQSMILVIDHLITYYQ 193  
 +++L + + + +++DE + LD SS +LQL+ Q Q ++I H I

Sbjct: 366 QKILQILMIMTKRNVLLIDEPLSGLLHBSVDLVVLQMLQRCQEKIQQTFLIISHQIDALA 425

Query: 194 DIMCH 198

D D+

Sbjct: 426 DFCDY 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4415> which encodes the amino acid sequence <SEQ ID 4416>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3093 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 120/455 (26%), Positives = 203/455 (44%), Gaps = 47/455 (10%)

Query: 1 MLSVSKIACTHGDHYLFDEV--TFSLNPQRILISGYSCQKSTLALLGLKE---SGK 56  
 M+S E+L T+ D ++ T + G+ I++ G SG GKST LL+G+ +GK

Sbjct: 21 MISAEQLVFTYTHDQKNFACQISTCTQIASGQFIVLQGPSGSGKSTFLKLIANGIIPDYAGK 80

Query: 57 GQVLLNLSLIEPS-----DVGFLEQNPDLQPCMDTVARHLYFILENLQIEPEQMD 107

+ L+ + + V +FCNP QF V HEL F EN ++ + +  
 Sbjct: 81 YERGLDQVADQAGRDSVETPSRSVASVFQNPASQFFYREVQVHSLVFPCEISQGLDAKVMK 140

Query: 108 RSEFVLAQVGLKGFQNRILYTLISOGKQRALATIFLSPKGLIILDEAFANLDQESASQL 167

R + N+ ++ LS G+QR+A+AT ++ +++ DE ANLD + +  
 Sbjct: 141 RLNTLAEDPAPAEALLNKDMFGLSGGQQRVAIATAINQSTNIMLFDEPTANLQAGIAAV 200

Query: 168 LQLVLVNYQANNQSMILVIDHLITYYQDMDHYFM-----LEKRLTRVNF-----DY 213

+ +A ++ +IV +H + Y D+ D++F+ L +LT N D  
 Sbjct: 201 KAYLTQLKAAGKT--ITVAERHLHYIMDLADNFFYFKNGRLTDKLTQNLALITBQRDM 259

Query: 214 MLNRLNVFELE-----KKSHTGDKLLSKDFQVLSKNKFIISYLDQFSLASGERLCLD 266

L RL++ +L+ + + H D L I+ V+ + A G +  
 Sbjct: 260 GLRLDLDLKPVLGKTESQHYRPDLSLCIEHLTVRAGSKILRCIBQLSFAVGSISGTT 319

Query: 267 GPSGVKSSSLFNGLLGLRTYKGIQFTHRKQIPISLFQNPVLDQFIPSTVYDIF--CVC 324

G +G+KS L + G+ KK + IP+S + + + V ++F V  
 Sbjct: 320 GSNGLGKSQLVYIAGT--LDDKRTATIKQGIPLSAKQRLEKTSIVLQEVSLQFAESVS 377

Query: 325 KDN-----KARDILSTINLMDKKQSPFQGLQCGQRIATQGISLASDGLLLLEPT 377

K+ N + +++E ++L + P LS G+Q+R+ I + L +D ++L DEP+  
 Sbjct: 378 KEVNLGHERHPRPTTEVIERLSLITLLERHPASLGGGQQRVMAASLLADKILIFDEPS 437

Query: 378 YQDQYHANMITTLISYCHKNHGVFTPSHDDL 412

G D + LL+ H VI SHD L

Sbjct: 438 SGLDLQMKALANLLMQ-LKQHKVVILISHDEL 471

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1439

A DNA sequence (GBSx1525) was identified in *S.agalactiae* <SEQ ID 4417> which encodes the amino acid sequence <SEQ ID 4418>. Analysis of this protein sequence reveals the following:

-1588-

Possible site: 42

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

5  
 INTEGRAL Likelihood = -11.62 Transmembrane 8 - 24 ( 1 - 30)  
 INTEGRAL Likelihood = -8.17 Transmembrane 145 - 161 ( 143 - 163)  
 INTEGRAL Likelihood = -6.32 Transmembrane 66 - 82 ( 62 - 84)  
 INTEGRAL Likelihood = -3.77 Transmembrane 112 - 128 ( 111 - 132)  
 INTEGRAL Likelihood = -2.66 Transmembrane 43 - 59 ( 43 - 59)

----- Final Results -----

10  
 bacterial membrane --- Certainty=0.5649(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15  
 >GP:CB13180 GB:Z99110 ykoB [Bacillus subtilis]  
 Identities = 68/177 (38%), Positives = 117/177 (65%), Gaps = 1/177 (0%)  
 Query: 5 LKDVLLIALLAVLVGVVYFGAGYISNAFVFFVGPIAHEVIYGINFVAGPMALYILRKPT 64  
 +K++++++V VVY + N GPIA+E IYGINF+ +A Y++RKP  
 20  
 Sbjct: 6 VKSEIVMSVISIVFAVYLLDFTHFGNVLGMPGPIAYEPTIYGINFVIAAYMIRKPGA 65  
 Query: 65 AIVBELAALIEVLIGSIYGPSVLVIGTLQGLSGELGFTLFRYHMYKLPAPLSAILTSI 124  
 A+V+E+AAL+E L+G GP V+VIG +QSLG+E F R+ Y LP +L+ +S+  
 25  
 Sbjct: 66 ALVSEIIALVECLLGNPSGPMVIVIGIVGSLGAEAVFLATRNKAYSLPVLMLAGSGSSV 125  
 Query: 125 FSPFAMFYANGLSAFSFSYNIMLIVRTVS-SIIFLLTNICDQLHRSGLVNAIGI 180  
 SF + + +G +A+S Y ++ML++R +S +++ LL K + L +GVLN +  
 Sbjct: 126 ASFYIDLFSYGYAAYSPGYLLIMLVIRLISGALLAGLLGKAVSGSLAYTVGIANMAL 182

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1440**

A DNA sequence (GBSx1526) was identified in *S.galactiae* <SEQ ID 4419> which encodes the amino acid sequence <SEQ ID 4420>. Analysis of this protein sequence reveals the following:

Possible site: 47

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

40  
 INTEGRAL Likelihood = -6.69 Transmembrane 65 - 81 ( 53 - 95)  
 INTEGRAL Likelihood = -6.37 Transmembrane 34 - 50 ( 31 - 54)  
 INTEGRAL Likelihood = -6.10 Transmembrane 176 - 192 ( 169 - 195)  
 INTEGRAL Likelihood = -3.66 Transmembrane 130 - 146 ( 130 - 151)  
 INTEGRAL Likelihood = -1.97 Transmembrane 3 - 19 ( 3 - 19)  
 INTEGRAL Likelihood = -0.90 Transmembrane 88 - 104 ( 88 - 104)

----- Final Results -----

45  
 bacterial membrane --- Certainty=0.3675(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9757> which encodes amino acid sequence <SEQ ID 9758> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

55 A related GBS gene <SEQ ID 8809> and protein <SEQ ID 8810> were also identified. Analysis of this protein sequence reveals the following:

-1589-

Lipop: Possible site: -1   Crend: 8  
 McG: Discrim Score:     -4.09  
 GVH: Signal Score (-7.5): -4.38  
 Possible site: 47

5     >>> Seems to have no N-terminal signal sequence

ALOM program    count: 6 value: -6.69 threshold: 0.0

INTEGRAL	Likelihood = -6.69	Transmembrane	65 - 81 ( 53 - 95)
INTEGRAL	Likelihood = -6.37	Transmembrane	34 - 50 ( 31 - 54)
INTEGRAL	Likelihood = -6.10	Transmembrane	176 - 192 ( 169 - 195)
INTEGRAL	Likelihood = -3.66	Transmembrane	130 - 146 ( 130 - 151)
INTEGRAL	Likelihood = -1.97	Transmembrane	3 - 19 ( 3 - 19)
INTEGRAL	Likelihood = -0.90	Transmembrane	88 - 104 ( 88 - 104)
PERIPHERAL	Likelihood = 5.30		158

modified ALOM score: 1.84

15     \*\*\* Reasoning Step: 3

      ----- Final Results -----

bacterial membrane	--- Certainty=0.3675(Affirmative) < succ>
bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25     Example 1441

A DNA sequence (GBSx1527) was identified in *S. agalactiae* <SEQ ID 4421> which encodes the amino acid sequence <SEQ ID 4422>. Analysis of this protein sequence reveals the following:

Possible site: 23

30     >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----

bacterial outside	--- Certainty=0.3000(Affirmative) < succ>
bacterial membrane	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 40     A related GBS gene <SEQ ID 8811> and protein <SEQ ID 8812> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1   Crend: 2  
 McG: Discrim Score:     6.01  
 GVH: Signal Score (-7.5): 0.45  
 Possible site: 23

45     >>> Seems to have a cleavable N-term signal seq.

ALOM program    count: 0 value: 10.66 threshold: 0.0

PERIPHERAL	Likelihood = 10.66	80
------------	--------------------	----

modified ALOM score: -2.63

50     \*\*\* Reasoning Step: 3

      ----- Final Results -----

bacterial outside	--- Certainty=0.3000(Affirmative) < succ>
bacterial membrane	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

-1590-

SEQ ID 4422 (GBS19) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 4; MW 24kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 6; MW 46.1kDa).

The GST-fusion protein was purified as shown in Figure 190, lane 10.

#### 5 Example 1442

A DNA sequence (GBSx1528) was identified in *S.agalactiae* <SEQ ID 4423> which encodes the amino acid sequence <SEQ ID 4424>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have a cleavable N-term signal seq.  
 10 ----- Final Results -----  
           bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>  
 15

A related GBS nucleic acid sequence <SEQ ID 8813> which encodes amino acid sequence <SEQ ID 8814> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1   Crend: 6  
 SRCLFG: 0  
 20 MoG: Length of UR: 23  
       Peak Value of UR: 2.61  
       Net Charge of CR: 3  
 MoG: Discrim Score: 9.08  
 GVH: Signal Score (-7.5): -0.76  
 25 Possible site: 22  
 >>> Seems to have a cleavable N-term signal seq.  
 Amino Acid Composition: calculated from 23  
 ALOM program count: 0 value: 5.14 threshold: 0.0  
 PERIPHERAL Likelihood = 5.14   365  
 30 modified ALOM score: -1.53  
 \*\*\* Reasoning Step: 3  
 Rule gpol  
 35 ----- Final Results -----  
           bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>  
 40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA34476 GB:X16457 precursor polypeptide (AA -26 to 632)  
           [Staphylococcus aureus]  
 45 Identities = 93/372 (25%), Positives = 160/372 (43%), Gaps = 46/372 (12%)  
 Query: 9 MKQFLKSAAILSLAVTAVSTSQPVGAIVGKDETKLRQQLGYIDSKSKKIDDERWGEKI 68  
           MKKQ + A L+ +A + + AIV KD +K + + K G + + + KI  
 Sbjct: 1 MKKQILSLGA-LAVASSLFTWKNKADAIVTKDYSK--BSRVNEKSKKGATVSDIYYWKI 56  
 50 Query: 69 YNYLSYELIEANENWINRSEPOEYRTITLSEPKDKIDSIEYVLINLS-----NIAKEDAHQ 124  
           + L + A + + ++ +P Y+ ++ + YL+ + K+  
 Sbjct: 57 IDSLAQFTGAIDLLNNTKYGDPIYKRAKDRIMTRVILGEDQYLLKKKIDRYELYKRWYKS 116  
 Query: 125 RNTLQSLDKYKSGIYNLQGVYNYIYQRISSAKHKFSQGVKIRYLDSTLFFPSWYDK 184  
           N ++ + K +YNL YN I+ + A +P+ V +I + L F  
 55 Sbjct: 117 SNKNNLNLTHPKNYLNTMNEYNDIFNSLKDAVYQPNKEVKIEIHQNVDLKQF----- 170  
 Query: 185 HLCNNENYKDNKDKFKFYIALLNEITRKARIGYQIVNNHKD-GEHKDEARI-LDILIRIDT 242



-1591-

D ++K KE L++EI Y KD GRH E LD+++ D  
 Sbjct: 171 -----DKDGEDKATKGVYDLVSEIDTLVVITYA-----DKDYGEHAKELRAKLDLILGDTD 221  
 Query: 243 FVSKDAPGYKIPHKRIAAKIIEIDLGIINDFFKNIGKDKP-SLEKIKDTPEHKKIYINST 301  
 K I N+RI ++I+DL+ II+DPF T +++P S+ K T+ + K +  
 Sbjct: 222 NPKH-----ITNERIKKEMIDDLNSIIDPFMETKQNRNFSITKYDPTKRNFKESSEN 274  
 Query: 302 EPTYSIBTNLPSNYKELKEKQIKKLEYGYK-KSSKTY--TSAHYALYSEIDANKELLQKV 358  
 +P N +E K K +K+ + +K K+ K Y T + E+ + L V  
 Sbjct: 275 KP-----NFDKLVETK-KAVKEADBSWKNKTVKRYEETVTKSPVVKEEKKVEEQLPKV 328  
 Query: 359 KIAKDNYNIEIKS 370  
 N E+K+  
 Sbjct: 329 ----GNQGEVKT 336

No corresponding DNA sequence was identified in *S. pyogenes*.

SEQ ID 8814 (GBS119) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 2; MW 84.3kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 5; 2 bands).

The GBS119-GST fusion product was purified (Figure 109A; see also Figure 201, lane 6) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 109B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1443

A DNA sequence (GBSx1529) was identified in *S. galactiae* <SEQ ID 4425> which encodes the amino acid sequence <SEQ ID 4426>. This protein is predicted to be s-adenosylmethionine synthetase (metK). Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3609 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07019 GB:AF001518 S-adenosylmethionine synthetase [Bacillus halodurans]  
 Identities = 266/390 (68%), Positives = 324/390 (82%), Gaps = 1/390 (0%)  
 Query: 4 RKLFTSESVSGHGPDKIADQISDAILDALILEQDPDAHVAAKTAVTSGSVHVFGEISTTAY 63  
 R+LFTSESV+EGHGPDKI DQISD+ILD IL+DPA VA ET+V TG V V GEI+T+ Y  
 Sbjct: 7 RRLFTSESVTEGHGPDKICDQISDILDEILKEDPNARVACETSVITGLVLVAGEITTSY 66  
 Query: 64 VDINRVVRNTIARIGYDKAEGFSAESVGHPSLVHSPDIAQGVNALEVR-GLSEQDP 122  
 VDI +VVR+TI IGY +A+YGF +E+ V S+ EQSPDIAQGVN+ALE R G +  
 Sbjct: 67 VDIKPVVRDTRINTGYTRAKYGFDSCTCAVLFTSDQSPDIAQGVNALEAREQQTDAE 126  
 Query: 123 LDIGAGDQGLMFGFVDETEPELMPLEISLHQLVKLITLHRSKGLTYLRPDIAKSVQTV 182  
 ++ IGAGDQGLMFG+A +ETPELMPLEISL+H+L +L+ RK L YLRPD K+QVTV  
 Sbjct: 127 TRATIGAGDQGLMFGYANNETPELMPLEISLHKLARLSKARKGELIPLRYPGKTVQTV 186  
 Query: 183 EYDENDQPIKVDAAVISTQHDENVINDOLEKDVIEKVINSEVIPSHYLDQTKFFINPTGR 242

-1592-

```

      EYDENDQ +R+D +VISTQH F VT +Q+ D+ + VI V+P +D++TK+FINPTGR
Sbjct: 187 EYDENDQSVRIDTIVISTQHHEVTLTQIESDLKQHVIRSVVPEELIDETKYFINPTGR 246

5   Query: 243 FVIGGPGQDSGLTGRKIIIVDTYGGYSRHGGGAFSGKDATKVDRSASVAARYIAKNIIVAD 302
      FVIGGPGQD+GLTGRKIIIVDTYGGY+RHGGGAFSGKD EKVDRS +YAARY+AKNIVAA
Sbjct: 247 FVIGGPGQDAGLTGRKIIIVDTYGGYARHGGGAFSGKDP TKVDRSAGYAARYVAKNIVAA 306

10  Query: 303 LAKKVEVQLAYAYGVAQPVSVRVDTTGTGVTIARADLEAAVRQIFDLRPAGINMLDLKRP 362
      LA K EVQLAYAYGVA+PVSV +DTPTG ++EA L VR+ FDLRPAGIN MLDL+RP
Sbjct: 307 LADKCEVQLAYAYGVAQPVSVISIDTPTGTQVSEAKLVKLVREHFDLRRPAGITIM/LDLRRP 366

15  Query: 363 IYRQTAAYGHMGRITIDLPWERVDKVAQLK 392
      IY+QTAAYGH GRITD++LFEW+ DK + L+
Sbjct: 367 IYKQTAAYGHMGRITIDLPWERVDKVAQLK 396

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4427> which encodes the amino acid sequence <SEQ ID 4428>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence

20  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3389(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 333/395 (84%), Positives = 361/395 (91%), Gaps = 1/395 (0%)

30  Query: 1 MSERKLPSTSESVSGHFDKIAQISDAILDAILBQDPAHVAABTAVYTGSHVHVGEST 60
      MSERKLPSTSESVSGHFDKIAQISDAILDAIL +DP+ARVAABT VYTGSHVHVGEST
Sbjct: 1 MSERKLPSTSESVSGHFDKIAQISDAILDAILAEDPEARVAABTCTVYTGSHVHVGEST 60

35  Query: 61 TAYVDINRVVNTIARIIGYDKAEYGFSAESVGVHPSLVBQSPDIAQGVNEALEVRGSLAQ 120
      TAY+DINRVVR+TIARIIGY +AEYGFSAESVGVHPSLVBQS DIAQGVNEA E R +
Sbjct: 61 TAYVDINRVVNTIARIIGYTEAYGFSAESVGVHPSLVBQSDIAQGVNEAFESREG-DT 119

40  Query: 121 DPLDLIGAGDQGLMFGFAVDETFPELMPLPISLAHQLVKKLTDLRKSGBELTYLRPDAKSV 180
      D L IGAGDQGLMFGFA++ETPELMPLPISLAHQLV++L +LRKSGB++LYLRPDAKSV
Sbjct: 120 DDLSHIGAGDQGLMFGFALINETPELMPLPISLAHQLVKRLAELRKSGBEISYLRPDAKSV 179

45  Query: 181 TVEYDENDQPIRVDAVVISTQHDENVNTQDHLKDVIEKVINEVIPSISYLDQTKFPINPT 240
      TVKYDE+D+P+RVD VVISTQHDPT TNDQ+ +DVIEKVI VIP+ YLDD TKFPINPT
Sbjct: 180 TVEYDEHDKPVRVDTVVISTQHDPEATNDQIRQDVIEKVIKAVIPADYLDLDDTKFPINPT 239

50  Query: 241 GPFVIGGPGQDSGLTGRKIIIVDTYGGYSRHGGGAFSGKDATKVDRSASVAARYIAKNIIV 300
      GPFVIGGPGQDSGLTGRKIIIVDTYGGYSRHGGGAFSGKDATKVDRSASVAARYIAKNI+VA
Sbjct: 240 GPFVIGGPGQDSGLTGRKIIIVDTYGGYSRHGGGAFSGKDATKVDRSASVAARYIAKNI+VA 299

55  Query: 301 ADLAKKVEVQLAYAYGVAQPVSVRVDTTGTGVTIARADLEAAVRQIFDLRPAGINMLDLK 360
      A L K EVQLAYAYGVAQPVSVRVDTTGT + EA LEAAVRQIFDLRPAGINIMMLDLK
Sbjct: 300 AGLVTKAEVQLAYAYGVAQPVSVRVDTTGTVEPVALEAAVRQVFDLRRPAGITIM/LDLK 359

      Query: 361 RPIYRQTAAYGHMGRITIDLPWERVDKVAQLKDI 395
      RPIY+QTAAYGHMGRITIDLPWER++KV AL + +
Sbjct: 360 RPIYKQTAAYGHMGRITIDLPWERLAKVDALVEAV 394

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1444

A DNA sequence (GBSx1530) was identified in *S. agalactiae* <SEQ ID 4429> which encodes the amino acid sequence <SEQ ID 4430>. This protein is predicted to be a transcriptional repressor of the biotin operon. Analysis of this protein sequence reveals the following:

```

5   Possible site: 24
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.16    Transmembrane 188 - 204 ( 188 - 204)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9755> which encodes amino acid sequence <SEQ ID 9756> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
      operon [Bacillus halodurans]
      Identities = 102/315 (32%), Positives = 169/315 (53%), Gaps = 18/315 (5%)

20  Query: 10  ILSKNNFISGETYANQLNISRTAIWGIKLTLEELGLEISVTNKGRLVSG-DILLPEQ 68
      +L+ ++P+SGE ++ + SRTA+WK I+ L + G E+E+V KGYR+V D + P
      Sbjct: 9  LMTAGDDPVSFGKISQAIGCSRTAVWKHIEELRKSQTYEVAQRKGYRIVRDPQIKPHD 68

25  Query: 69  LE-----QEIGIKVSLNNNSASTOLDAKMGIESKLTTPHFLFAPNKKAKGRFDRPFPTS 123
      ++ + G +++ ++ASTQ A + K H+ LA Q KGR R +++
      Sbjct: 69  IQVVLTERPGRBITYLESTASTQTVALKLAQBGAKGSHIVLANEQTSKGKMGKRWYSP 128

30  Query: 124  NQGGIYMSLLQPNVPIDIKPYTMVASSAVKAISRLTGITPEIKWVNDIYLDNKKIAG 183
      I MS++ +P +P + T++ A + V+AI TG+ +IKW ND+ +D KKI G
      Sbjct: 129  PGSSISMSIIFRPQLPPQAPQLTLTAVALVRAIKETTGLOSDIKWPNLLDIDGKKIAG 188

35  Query: 184  ILTEAIASVSGSLVTNVIIGLGINFYIKE--FPRALTKRAGSLFTEQ-PTITENQLITEI 240
      ILTE A +8 V +VI G+GIN +E F + K A SL ++ I R LI I
      Sbjct: 189  ILTEMQADQDS--VHSVVIQIGINVNHQEEAFABEIRKIATSIAIKKGEPIQRAPLIAI 246

      Query: 241  W---NLFFNIPLEDHLK---VYREKSLVLDKTVSPMDGQTMYSKKAIDITDGGVLVEL 293
      LF+++ L+ ++ ++ + + + G A ITD G L++E
      Sbjct: 247  LQNIELFYDLYLQHGFRIKPLWEAHAISIGKIRARMINDVKFGVARGITDDGVLLED 306

40  Query: 294  DDQQLKTLRSGEISL 308
      DDG+L ++ S +I +
      Sbjct: 307  DDGLHSIYSADIEI 321

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4431> which encodes the amino acid sequence <SEQ ID 4432>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 34
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.49    Transmembrane 194 - 210 ( 194 - 211)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1595 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
      operon [Bacillus halodurans]
      Identities = 98/315 (31%), Positives = 165/315 (52%), Gaps = 18/315 (5%)

```

-1594-

Query: 10 LLSQTDDFVSGRYLADQLSISRTSVWKSIKSLBNQGIQIDSLKHKGYRMVQG-DILLPKT 68  
 LL+ DDFVSGE ++ + SRT+VWK I+ L G +++++ KGYR+V+ D + P  
 5 Sbjct: 9 LLTAGDDFVSGREKISQATGCSRTAVMKHIEELRKSQGYEAVQNGSYAIVKRPDQIKPHD 68

Query: 69 I-----SQGLGMFVTTYPHSQSTQLDAKQGIEAMNSAPRLYLAPSQBAANGRLDRQFFSA 123  
 I ++ G +TY + STQ A + + + La Q + RGR+ R ++S  
 10 Sbjct: 69 IQVVLSTYRPGREITTLSTASTQITVALKLAQGEAKGHIVLANEQTSKGKMGKRGWVSP 128

Query: 124 STGGIYMSMYLKPNVPYADMPFYTMVASSIVKASRLTGIDTEIKWNDIYLGNHVKAG 183  
 I MS+ +P +P P T++ A +IV+AI TG+D++IKW ND+ + K+ G  
 10 Sbjct: 125 PGSSISMSITTFPQLFPQKAPQLTLTAVAVRAIKETTLGDSDIKWNDLLIDGKKIVG 188

Query: 184 ILTEAITSVETGLITTVIIGVGLNFPVTD--FPEIAQKAGSLFTEK-PTITRNDLIIDI 240  
 ILTE + + VI G+G+N + F E I + A SL +K I R LI I  
 15 Sbjct: 185 ILTE--MQADQDSVHSVTCIGINVMHQBFAFEETIKIATSLAIKKEPIQAPLITAI 246

Query: 241 WK-----LFLSIPVKDHVKVYKESLVLNKQVTFIENSQEKRAITADLTQGHILIVQF 293  
 K L+L +++ ++ + K++ + K +A +TD G L+++  
 20 Sbjct: 247 LKNIELFDIYLQHGFSRIKPLMEAHALISGKRIRARMINDVKFGVAGKITDGGVLLLED 306

Query: 294 ENGDLQTLRSGEISL 308  
 ++G L ++ S +I +  
 25 Sbjct: 307 DDGKLHSIYSADITEI 321

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/311 (61%), Positives = 257/311 (82%)

Query: 1 MKTYEKIVQLSKNNFISGETMANQLNISRTAIWNGIKTLRELGLEIBSVNKGRLVS 60  
 MKT EKIVQL+L+ ++F+SQE +A+QL+ISRT++WK IK+LE G++I+S+ +KGYR+V  
 30 Sbjct: 1 MKTSEKIVQLSQDQDDFVSGRYLADQLSISRTSVWKSIKSLBNQGIQIDSLKHKGYRMVQ 60

Query: 61 GDILLPEQLGQIGIKVSLNNNSASTQLDAKMGIESKLATPHLFLAPNKKAKGRDRPF 120  
 GDILL+ + Q +G+ V+ +S STQLDAK GIE+ P L+LAP+Q+ AKGR DR F  
 35 Sbjct: 61 GDILLPKTISQGLGMFVTTYPHSQSTQLDAKQGEAENSAPRLYLAPSQBAANGRLDRQF 120

Query: 121 FTSNQGGIYMSLLQPNVPNDIKPYTMVASSAVKASIRLTGITEIKWVNDIYLDNKK 180  
 F+++ GGIYMS+ L+PNVP D+ FYT+MVASS VKAISRLTGI EIKWVNDIYL N K  
 40 Sbjct: 121 PSASTGGIYMSMYLKPNVPYADMPFYTMVASSIVKASRLTGIDTEIKWVNDIYLGNH 180

Query: 181 IAGILTEIASVESGLVTVNIIGLGINFYIEPPRALTKRAGSLFTEQPTITRMQLITEI 240  
 +AGILTEAI SVE+GIL+T+VIIG+G+NF++ +FP A+ ++AGSLFTE+PTITRM LI +I  
 45 Sbjct: 181 VAGILTEAITSVETGLITDVIIIGVGLNFPVTDFFPEIAQKAGSLFTEKPTITRNDLIIDI 240

Query: 241 WNLFFNIPLDHLKLVYKESLVLDRTVSFMQGTMYSGKADITDKGYLVVLEDDGQKLT 300  
 W LF +IP++DH+KVY+EKSHWL++ V+F++ AID+DH+G+L+V+ ++G L+T  
 50 Sbjct: 241 WKLFLSIPVKDHVKVYKESLVLNKQVTFIENSQEKRAITADLTQGHILIVQFENGDLQT 300

Query: 301 LRSGRISLSSW 311  
 LRSGRISLSSW  
 50 Sbjct: 301 LRSGRISLSSW 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 55 Example 1445

A DNA sequence (GBSx1531) was identified in *S. agalactiae* <SEQ ID 4433> which encodes the amino acid sequence <SEQ ID 4434>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have an uncleavable N-term signal seq  
 60 INTEGRAL Likelihood = -2.76 Transmembrane 3 - 19 ( 3 - 20)  
 ----- Final Results -----

-1595-

```

bacterial membrane --- Certainty=0.2105 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1446

10 A DNA sequence (GBSx1532) was identified in *S.agalactiae* <SEQ ID 4435> which encodes the amino acid sequence <SEQ ID 4436>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -2.28 Transmembrane 24 - 40 ( 24 - 40)

----- Final Results -----
bacterial membrane --- Certainty=0.1914 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

20 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4437> which encodes the amino acid sequence <SEQ ID 4438>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -1.91 Transmembrane 58 - 74 ( 58 - 75)

----- Final Results -----
bacterial membrane --- Certainty=0.1765 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 37/67 (55%), Positives = 54/67 (80%), Gaps = 3/67 (4%)

Query: 1  MTKRQFIWALLCSFETYFFNQSVMDGSIWIFAIWGVLLLRDLQKVYAIKFKTELK-- 58
          MT RQF+FWA +C+PETYFFN ++ G+++FA+FWG+LL RDL++V+ I++ TK ++K
Sbjct: 36  MTKRQFLWAFVCAFPETYFFNDLLSGNYLFAFWGILLFRDLRFVHTINQLTKTLKTA 95

Query: 59  -STKKKD 64
          S KKKD
Sbjct: 96  NSPKKKD 102

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1447

A DNA sequence (GBSx1533) was identified in *S.agalactiae* <SEQ ID 4439> which encodes the amino acid sequence <SEQ ID 4440>. This protein is predicted to be DNA polymerase III, gamma subunit (dnaZX). Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have no N-terminal signal sequence

```

-1596-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1567 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4441> which encodes the amino acid sequence <SEQ ID 4442>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.59 Transmembrane 232 - 248 ( 232 - 249)

## ----- Final Results -----

bacterial membrane --- Certainty=0.1235 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 408/558 (73%), Positives = 473/558 (84%), Gaps = 6/558 (1%)

```

Query: 1  MYCALRYKYSQTFDEMVGQSVISTTLKQAVSSKKISHAYLFSGPRGTGKTSAAKIFAKA 60
      MYCALRYKYSQTFDEMVGQSVISTTLKQAV S KISHAYLFSGPRGTGKTSAAKIFAKA
Sbjct: 1  MYCALRYKYSQTFDEMVGQSVISTTLKQAVSGKISHAYLFSGPRGTGKTSAAKIFAKA 60

Query: 61  MNCPNQINGEPCNQCDCIDRTINGSLDEVVIDEADSNNGVDEIRDIDKSTYAPSRATYK 120
      MNCPNQ++GEPCN CDI CRDITNGSLDEVVIDEADSNNGVDEIRDIDKSTYAPSRATYK
Sbjct: 61  MNCPNQVDSGPCNQCDCIDRTINGSLDEVVIDEADSNNGVDEIRDIDKSTYAPSRATYK 120

Query: 121  VYIIDEVHMLSTGAFHALLKXLEPTENVVFILATTELHKIPATILSRVORFEFKAIKLL 180
      VYIIDEVHMLSTGAFHALLKXLEPTENVVFILATTELHKIPATILSRVORFEFKAIK
Sbjct: 121  VYIIDEVHMLSTGAFHALLKXLEPTENVVFILATTELHKIPATILSRVORFEFKAIKQK 180

Query: 181  AIRDHLAQITLDEKAISYDELDAITVARRAEGGMRDALSTLDQALSKADNHISLDVAEEI 240
      AIR+HLA +LDKE I+Y+DAL L+ARRAEGGMRDALSTLDQALSL+ DN +++ +AEEI
Sbjct: 181  AIRKHLAVLDKEGIAVEVDALNIARRAEGGMRDALSTLDQALSLSPNQVAIAIAEEI 240

Query: 241  TGSISLSAIDYVSNILAHDTTELAKLEIVDFGSKMSRFATDLLVLELLVQVQAGE 300
      TGSIS+ A+ DYV + +T+ALA LR I+DSGSKMSRFATDLL YLELLLV+AG+
Sbjct: 241  TGSISLALGDYVRYVSQEQATQALAALSTIYDSGSKMSRFATDLLTYLELLLVKAGD 300

Query: 301  DSHSSTDFIANLVKQDILFEMIDKYSVLPEIKNGSHFKVYAEEMTIQLSGMEVENS- 359
      + S F NL++ D +F+MI VTS LPRIK G+HP+YAEEMTIQI++ + S
Sbjct: 301  NQRQSAVFDNLSESDRIPQMIVTVTSRLPEIKGTHPRIYAEEMTIQIAQKEQLSQV 360

Query: 360  NIPADVIAELSLRLRELKSLKNEMQL-SPADQSSSTCKVKVNNKFTTFPKVDRTKILTM 418
      N+ ++ +S++L+ EL LK ++SQL SR D + + K K KT +++VDR IL IM
Sbjct: 361  NLSGELISEIETLKLNLQALQQLQQLQSRPDSLARSDKTK--PKTTSYRVDVITLKLIM 418

Query: 419  EETVDSQSRSEFYLEALKSNWNEILDNITACDRALLMGSEFVLANSENAILAFDAFNAE 478
      EETV +S0+SR+YL+ALK+ANNHILDNIT+ACDRALLMGSEFVLANSENAILAF+AFNAE
Sbjct: 419  EETVNSQSRQYLDALNANWNEILDNISAQDRALLMGSEFVLANSENAILAFEAFAFAE 478

Query: 479  QAMKRITLDNIFGNIMSKAAGFSENLIVAPENDFNQIRSDFAKKMKQAK--TETEPEVNI 536
      Q N R +LND+FGNIMSKAAGFSENLIVAPR DF IR +FA++MK+QK + E EV
Sbjct: 479  QVNSRNLLNMFGNIMSKAAGFSENLIVAPRTDFQHIRKFAQMKQSKQDSVQBEQVAL 538

Query: 537  CIPEDFSYLAERIAIVED 554
      IPE F +L ++I ++D
Sbjct: 539  DIPEGDFLLDKINTIDD 556

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1597-

**Example 1448**

A DNA sequence (GBSx1534) was identified in *S. agalactiae* <SEQ ID 4443> which encodes the amino acid sequence <SEQ ID 4444>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence (or aa 1-19)

----- Final Results -----

bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06927 GB:AP001518 unknown conserved protein [Bacillus halodurans]  
 Identities = 67/143 (46%), Positives = 96/143 (66%)

Query: 8 KNYQLLLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLDGEEILGPFQGGV 67  
 E Y L+ Q AL E++A+ANL+NASA+L L + GFYL EL+LGPFQGG  
 Sbjct: 13 EKYSLVTRQLAALLGEGSDAIAIANANASALYHFLFEEVNVVGFYLIKGEELVLGPFQGLP 72

Query: 68 SCVHITLGRGVCGESAQTAKTILVDDVTKHANYISCDKAMSEIVVPMFGKLLGVLDL 127  
 +CV I +G+GVCG +A+ +T+ V+DV + +I+CD+ + SRIV+P+P+NG I GVLID+  
 Sbjct: 73 ACVRIPIGRGVCGTAAKEBQTVRVDDVHQFPGHICDAASRSEIVIFLPQNGVLYGVLDI 132

Query: 128 DSSLVADYDEIDQGYLEKFGVIL 150  
 DS + + E +Q LE FV +L  
 Sbjct: 133 DSPSLNRFSEESQALLSFVDVL 155

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4445> which encodes the amino acid sequence <SEQ ID 4446>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1753 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 122/164 (74%), Positives = 144/164 (87%)

Query: 1 MNKSKKIENYQLLLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLDGEEIL 60  
 MNKSKKIE YQL++ QA+ LF++E+NALANLSNAGA+LN LPSVFTGFYLDG+ELIL  
 Sbjct: 1 MNKSKKIRYQYIMDAQAKLFANESNALANLSNALLNMLPNSVFTGFYLDGQELIL 60

Query: 61 GPFQGVSCVHITLGRGVCGESAQTAKTILVDDVTKHANYISCDKAMSEIVVPMFGK 120  
 GPFQGVSCVHI LGRGVCGESAQ+ +T+I+DV +HANYISCD+ AMSEIVVPM K G  
 Sbjct: 61 GPFQGVSCVHITLGRGVCGESAGSKRTI+INDVKQHANYISCDRAAMSEIVVPMVKGH 120

Query: 121 LIGVLDDSSLVADYDEIDQGYLEKFGVILVHETIWNLMFGVE 164  
 L+GVLDLDDSSLVADYDE+DQGYLE FV + +E T+ +MFGV+  
 Sbjct: 121 LIGVLDDSSLVADYDEVDQGYLEAFVDLPLEKTPTPTFMFGVK 164

SEQ ID 4444 (GBS282) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 9; MW 19.8kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 6; MW 44.8kDa) and in Figure 63 (lane 7; MW 47kDa).

-1598-

The GBS282-GST fusion product was purified (Figure 211, lane 4; see also Figure 225, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 269), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1449

A DNA sequence (GBSx1535) was identified in *S. agalactiae* <SEQ ID 4447> which encodes the amino acid sequence <SEQ ID 4448>. This protein is predicted to be uridine kinase (udk). Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB14675 GB:Z99117 uridine kinase [Bacillus subtilis]
Identities = 133/207 (64%), Positives = 167/207 (80%)

Query: 1 MRKKPIIIGVTGGGGGKTSVSRILSNFPDQKTIHEDSYKQSHLTFBERVKNTYD 60
      M K P++IG+ GGSG GKTSV+R+I F I MI+ D YYKQSHL FEER+ TNYD
Sbjct: 1 MGINPVPVIGIAGGGGGKTSVT+R+IYQFKGHSILMIQQQLYYKQSHLFFERLNTYD 60

Query: 61 HPLAFDTNLMIEQLNELIEGRFVDIPYDYTKHRSRDTIRQEPQDVIIIEGILVLESDQR 120
      HPLAFD + +IE + +L+ RP++ P+YDY HTRS+ T+ EP+DVII+EGILVLESD+R
Sbjct: 61 HPLAFDNDYLIIEHIQDLNLYRPIEKFTYDYKLHRSERTVHVEPQDVIIIEGILVLESDQR 120

Query: 121 LRDLMDIKLVYDITDDDIRIIRIKRMEERDRSLDSIIBQYTEVVKPMYHQFIEPTKRYA 180
      LRDLMDIKL+VYTD D+RIIRRI RD+ ER RS+DS+IBQY VV+PV++QF+EPKRYA
Sbjct: 121 LRDLMDIKLYVYTDADLRRIIRIRMDINERGRSIDSIVBQYVSVVRPMHQFVEPTKRYA 180

Query: 181 DIVIEGVSNIVAIDLINTKVASIIME 207
      DI+IPEG N VAIDL+ TK+ +IL +
Sbjct: 181 DIIIEGGQNHVADLNVTKIQITLEQ 207

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4449> which encodes the amino acid sequence <SEQ ID 4450>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9151> which encodes the amino acid sequence <SEQ ID 9152>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```



-1599-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/207 (83%), Positives = 193/207 (92%)

5 Query: 1 MRKKPIIIGVTGSGGGKTSVSRAILSNFPDQKITMIHDSYKQDQSHLTFEERVKITNYD 60  
M KKPIIIGVTGSGGGKTSVSRAIL +FP+ +I MI+HDSYKQDQSH++FEERVKITNYD  
Sbjct: 5 NLCKPIIIGVTGSGGGKTSVSRAILDSFPNARIAMIQHSYKQDQSHMFEEERVKITNYD 64

10 Query: 61 HPLAFDTNLMIEQLNELIEGRPVDPIDYDYTKHTRSDRTIRQRPQDVIIVRGILVIRDQR 120  
HPLAFDT+ MI+QL EL+ GRPVDPID+YDY KHTRS+ T RQ+PQDVIIVRGILVLEDR+R  
Sbjct: 65 HPLAFDTDFMIQQKLKELLAGRPVDIPIDYDYKHTRSNTTFRQDQDVIIVRGILVLEDER 124

15 Query: 121 LRDLMDIKLFPVTDDDIRIIRRIKRDNERDRSLDSIEIQTEVVVKPMYHQFIEPTKRYA 180  
LRDLMDIKLFPVTDDDIRIIRRIKRDNR ER RSL+SII+QYT VVKPMYHQFIEP+KRYA  
Sbjct: 125 LRDLMDIKLFPVTDDDIRIIRRIKRDNRGRSLSDIIDQYTSVVKPMYHQFIEPSKRYA 184

Query: 181 DIVIPRGVSNVAIDLINTKVASILNE 207  
DIVIPRGVSN+VAID+IN+K+ASIL E  
20 Sbjct: 185 DIVIPRGVSNVAIDVINSKASILGE 211

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1450**

- 25 A DNA sequence (GBSx1536) was identified in *S.agalactiae* <SEQ ID 4451> which encodes the amino acid sequence <SEQ ID 4452>. Analysis of this protein sequence reveals the following:

Possible site: 51  
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.5083(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CA512572 GB:Z99108 similar to RNA helicase [Bacillus subtilis]  
Identities = 140/343 (40%), Positives = 202/343 (58%), Gaps = 9/343 (2%)

40 Query: 10 QQKLTRQRQFDLTDIQNKLPQPIIDGNILGISPTGTGKTLAYFLPTLLKLPK-KSQQL 68  
Q+ F T+Q+ Q I DG+++ SPTGTGKTLAY P L+++P+ K Q  
Sbjct: 16 QENWASGQKFPQPTPVQEQVQLIMDGKVIASPTGTGKTLAYALVLEIRIKPQKHQQA 75

Query: 69 LILAPNSELAAQIPDVTKEWAEPLGLTAQLFLSGSSQKRIERLKKGPILLIGTAGRVFE 128  
+ILAP+ EL QIF V++W L A + G++ K+Q+H+LKK P I+GT GRVFE  
45 Sbjct: 76 VILAPSRKELVMQIPQVIGQWAGSELRAASLIGGANVKKQVRLKKHPHILVGTGKRVFE 135

Query: 129 LVKLAKIKMMNINTIVLDEFDELLGDSQYHFVDNIINRVPDQMIYISATNKLDS--- 185  
L+K KK+KM + TIVLE D+L+ + II RD+Q++ SAT K +  
50 Sbjct: 136 LIAKALKMHEVKITVLEFDITQLVLEPEREIMRQIIKTTLADRQLLCPSATLKKETEDVL 195

Query: 186 -KLADNTITIDLSNQKLOT--IKHYIITVDKRRITDILLKPSNIPDFRGLVFFNSISDLQ 242  
+LA + + K + +KH Y+ D+K++ LL+K S + + LVF + +L  
Sbjct: 196 RELAQREVLKVRQSKAEAGKVKHQYLLI CDQKDKVLLKQLKSLRLSGMALVFVRDITGLS 255

55 Query: 243 ACEERLQNRASAVSLASDINIKPKRVILEKFNHDSILLLOTDLVARGDIDNLEYVIN 302  
E+L ++ R I+ F++ + LLL TD+ ARG+DI+NL YVI+  
Sbjct: 256 VYAEKLAYHVELSVGLSEAKKMERAKIATPDEGEFPELLATDIAARGLDIENLPIYVIH 315

Query: 303 FDIARDKETTTHRSGRTRMGKSGCVITFVTHKRLKQLKRYA 345  
DI D++ Y HRSGRTRG QKSG V++ VT ER K LK A  
60 Sbjct: 316 ADIP-DEGYSVHRSGRTRAGKEGNVLSLVTKLEBSK-LKQMA 356

-1600-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4453> which encodes the amino acid sequence <SEQ ID 4454>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3847 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 273/358 (76%), Positives = 312/358 (86%)

15 Query: 1 MITKFPDQWQDKLTQRFQDDELITIQNKLPQPTDGNILGISPTGTGKTILAYLFPTLLKL 60
    MITKFP QWQ+KL Q F LT IQ + FQPI DG N LGISPTGTGKTILAY+FP LL L
    Sbjct: 12 MITKFPQWQEKLDQVAFHLLTPIQEQAFQPIVDGKKNFLGISPTGTGKTILAYVFPMLAL 71

20 Query: 61 QPKKSQQLLILAPNSELAQIFVITKWEAEPLGLTAQLFLSGSSQKRQIERLKKGPPEILI 120
    PKKSQQLLILAPN+ELAQIF+VTK+WA+PLGLTAQLF+SG+SQKRQIERLKKGPPEILI
    Sbjct: 72 TPKKSQQLLILAPNTELAQIFEVTKWQAFLGLTAQLFISGTSQKRQIERLKKGPPEILI 131

25 Query: 121 GTAGRVFELVKLKKIKMMNINTIVLDEFDELLGDSQVHFVDNIINRVPRDQMIYISATN 180
    GT GR+FEL+KLKKIKMM++NTIVLDE+DELLGDSQY FV I + VPD QM+Y+SATN
    Sbjct: 132 GTPGRIFELIKLKKIKMMNNTIVLDEYDELLGDSQVDFVQKISHYVPRCHQVMYMSATN 191

30 Query: 181 KLDSNKLADNTITIDLENQKLDITIKHYITVDKRRRTDLLRKFNSIPDFRGVLFNFSLS 240
    K+D + LA NT IDLS Q D I+H+Y+ VDKRRRTDLLRKF+NIPI R LVFNFSLS
    Sbjct: 192 KVDQTS LAPNTFCIDLEQNDIAIQHFYLMVDKRRRTDLLRKFNTIPHFRLVFNFSLS 251

35 Query: 241 LGACEERLQFNRAVASVLSADINIKFKFVILEKFNHDSILLGTDLVARGIDIDNLEIV 300
    LGA EERLQ-N A+AVSVLSADIN+KFRK ILEKFK+H +SLLL TDLVARGIDIDNL+YV
    Sbjct: 252 LGATEERLQNGNAAVASVLSADINVKFRKTILEKFKSHQLSLLATDLVARGIDIDNLEIV 311

40 Query: 301 INFDIARDKETHTHRSGRTERMGKSGCVITTVTHKEELQQLKKYATVTVLHNMQLH 358
    I+FD+ARDKE YTHR+GRTGRMGK G VITTV+H E+L+LKK+A V+E+ L NQ+LH
    Sbjct: 312 IHFDVARDKENYTHRAGRTGRMGKSGVITTVSHFEDLKLKKPAKVSISLQNGQLH 369

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1451

A DNA sequence (GBSx1537) was identified in *S.agalactiae* <SEQ ID 4455> which encodes the amino acid sequence <SEQ ID 4456>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL Likelihood = -1.38 Transmembrane 15 - 31 ( 13 - 31)

----- Final Results -----
      bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1601-

## Example 1452

A DNA sequence (GBSx1538) was identified in *S.agalactiae* <SEQ ID 4457> which encodes the amino acid sequence <SEQ ID 4458>. This protein is predicted to be peptidoglycan GlcNAc deacetylase. Analysis of this protein sequence reveals the following:

```

5      Possible site: 28
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -8.92    Transmembrane    4 - 20 ( 1 - 26)

10     ----- Final Results -----
      bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15     >GP:CA96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase
      [Streptococcus pneumoniae]
      Identities = 133/431 (30%), Positives = 228/431 (52%), Gaps = 20/431 (4%)

20     Query: 5   IIGIFSLII IAILAWQGFSLFKH--BIKQQAVVEKEIRIAEKTVVVKRQKTERVLF 62
      +IGI ++ I + + F + K E K++ EK+ ++E + RQ V+
      Sbjct: 21   LIGLAIISICLGGFI AFKIQKQSFQKIESLKEKDDQLSBCGKQKHFRCQAEVIA 80

      Query: 63   EPGKYDKLSADILKWNQKSFHKKFPYDNYIILRPQLADSNFANVKLSIQILYQKEK 122
      P +K +S+ NQ + + DN Q +S V ++ +Y
      Sbjct: 81   YPLQGEKVISVRELINQDVKDKLESKINLVFYITEQ--EESGLKGVVNRNVTQIYDLVA 139

      Query: 123  GSMFQKSSRLRLTYLTDQNKPPFLDELLAHNISGFKALENIAPGTQLK--EHDNSKEF 180
      + + L L ++ +PF LD+L + + + + + K E D +++
      Sbjct: 140  FKIRETEKTSLGKVLHTEDGQFFTLQDLQSLDASKAKQLIKELTSTFEDKIKTEQDQSQBI 199

30     Query: 181  LKTGRVTD---GLDVKDGKLI-----NDLKLPLDKLYNVIDESYLKSSDLDLVS 227
      +K D D KD ++I+ ++ LP+ ++VI SYL D L
      Sbjct: 200  VKNFSQDLQSAWNFYDKDSQIILYSPVVENLEETALPVSFAFFDVIQSYYLLEKDAALYQ 259

35     Query: 228  NLKAKAPR--VALTFDDGPNKHTPKALSIKRYNAKATFFVMQGSAGVHTDILQRMHAE 285
      + K + VALTFDDGPN TTP+ LE L +Y+ KATFFV+G++ G+ D++R++ +E
      Sbjct: 260  SYFDKKHQKVVALTFDDGPNPATTPQVLSTLAKYDIKATFFVLGKNSVGNEDLVKRIKSE 319

40     Query: 286  GHEIGNHTWDHPNLTLPARKIKSEIHKTNDLIMKATGQKFPVYLAPPYGATNAVTKVTVG 345
      GH +GNH+W HP L+L ++ K+I T D+ K G ++RPFYGA ++
      Sbjct: 320  GHVVGNIHSWHPILSGLSLEDAKQITTDTEVLTKVLGSSSLKMFPPYGAITDNRSLD 379

      Query: 346  LKEMLSVDTEDWKBQNTQAMWNIKQLRPGVILMHIDIHQITDALTPTIMDYLTGQY 405
      L ++W VD+ DWK+ N ++T I+ Q+ G ++LMDIH T++ALP +++YL QGY
      Sbjct: 380  LSPFMGDVDSLWKSKEASILTETIQGVANGSIVLMDHISPTVNALFRVETYLKNQGY 439

      Query: 406  YFVTVGELYST 416
      FVT+ E+ +T
      Sbjct: 440  TFFVTIPEMLNT 450
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4459> which encodes the amino acid sequence <SEQ ID 4460>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 22
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -12.58    Transmembrane    6 - 22 ( 1 - 27)

10     ----- Final Results -----
      bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

-1602-

The protein has homology with the following sequences in the databases:

IGS:AJ251472 peptidoglycan GlcNAc deacetylase [Strep... 239 4e-62

>GP:CAB96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase  
[Streptococcus pneumoniae]

Identities = 136/438 (31%), Positives = 230/438 (52%), Gaps = 23/438 (5%)

Query: 3 KLVNVLGKLLSILMLSLAI----VFINKWKLNEDSQRIVLARKKNTSDLVIAKVKHKK 58  
K +L+ L+ IL +S+ + + + + Q+I +K+K+ +H ++

Sbjct: 13 KTHVLLALIGLILAIISICLLGGPIAFKIYQKSKFSQKIESLKKEKDDQLSEGNQKEHFRQ 72

Query: 59 DQKDYFFSPK--QADFFVNLPLVSLYKKNSDKELLIVRPKIQSSHLRSVNTLTISK 116  
Q + + P++ + + + + K S L+ + + S L+ V ++K

Sbjct: 73 GQAEVIATYPLQGEKVISVRELINQDVKDKLESKDLVFTYTTQESSLKGKGVNRNVTK 132

Query: 117 IFYQKQFHLAKKSEKIVSTHVTDLDPQVQKDLVSGHL---ERIQEVEKKYCPDAGFN 173  
+T F + + + + H+T+D +PF + L S E+ +E+ D

Sbjct: 133 QIYDLVAFKIBETKTSIGKVHLTEDQGPFTLDQLFSDASKAKBQILKELTSPFIEDKKIE 192

Query: 174 SKYNGLEKESN-----LLSDGFVEKSKNLIID-----KKLTIPLTITLFDVINPDPLAN 222  
D+ + ++ S L + + F+ K +I + + + +P+ + FDVI +L

Sbjct: 193 QDQSEQIVKNFSDQQLSANNFQYKQSQIILYPSFVVENLEIALPVSAPFDVIQSSYLE 252

Query: 223 SDRAYVDNYRTYKEQHKKIALVTFDDGDPPTTPOVLDILAKYQAKGTFFMIGSKVAVN 282  
D A Y +Y K K Q K-VALTFDGDP+P TTPQVL+ LAKY K TFF+G V N

Sbjct: 253 KDAALYQSYFDKKHQ---KVALITYDDGPNPATTPQVLETLAKYDIKATFPVLKKNVSGN 309

Query: 283 ENLTKRVSDAGHIEAHNTWHDHFNLTLSVSEIQHVNMTQALEKACGKKRYLRPPYGA 342  
E+L KR+ GH + NH+W HP L+ L\$+ E + Q+ T + K G + +RPPYGA

Sbjct: 310 EDLVKRIKSBGVVGNHSHNHPILSQLSLDEAKKQITDTEVLTIKVLGSSSKLMRPPYGA 369

Query: 343 TNATVQSSGLTQMLNTVD+RWENHSTDGINVNQQLQPGGVVLAMHDHQTITNALPT 402  
++ S L+ ++N VD+ DW++ + I+T ++Q+ G +VLEHDH T+ALP

Sbjct: 370 ITDDIRNSDLDSFIMQDVSLSLWKSINSEILTEIQHVANGSIVLAMDHISPTVALPR 429

Query: 403 VMEYLKAGBVCVTISEL 420  
V+EYLK +GY VT+ E+

Sbjct: 430 VIEYLKNQGYTPVTIPEM 447

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 169/420 (40%), Positives = 259/420 (61%), Gaps = 12/420 (2%)

Query: 4 LIIGISPLIIILAILANQGSFLLGKHEIKLQAVVEKEIRIAKTVVWVKRQKTER--VLF 61  
+++G+ S+++++ LA + K E + + EK+ ++ ++ VK K ++

Sbjct: 7 ILVGLSILMLS-LAIVFINRWKLNEDSQRIVLARKKNTSDLVIAKVKHKIKKQDKYFF 65

Query: 62 LEPKDYKSLSDILKWNQKSPEHKKFYDNYIILRPQLADSNFANWKLKLSIYQLYQKE 121  
F D L S K K D + I++RP+L S+ +V L+I +I+YK+

Sbjct: 66 FSPKIQADFFVDNLP---VSLYKKNSDKELLIVRPKIQSSHLRSVNTLTISKIYQK 122

Query: 122 KGSMPKQSRLLRNTYLLDQNKPPDELLAHNISGFKALIENTAPCTQLKEDENKEFL 181  
+ +KS +++ TY + + KFF++ +L++ ++ + + E P N

Sbjct: 123 FPHLAKKSEKIVSTHVTDLDPQVQKDLVSGHLERIQEVEKKYCPDAGFNMDKVKLKE 182

Query: 182 NGRVTDGLDVKDGKLLIND-LKPLDKLKNVIDESYLSKSLDLVSNL--KAKAPR-- 235  
+ +DG +VK G LI + L +PL L+VI+ +L +SD N K + P+

Sbjct: 183 SNLLSDGFVEKSGNLIIDFKKLTIPLTITLFDVINPDPLANSIDRAYINRTYKQHPKGL 242

Query: 236 VALTFDDGKNTKTPKALEILKYNKAKATFFVQSGAVGHTDILQGMHAKSHHIGNTWD 295  
VALTFDDGP+ TTP+ L+L +Y AK TFF++G V + + + +R+ GHRI RHTWD

Sbjct: 243 VALTFDDGPDPTTTPQVLDILAKYQAKSTFFMIGSKVNNENLTKRVSDAGHIEAHNTWD 302

Query: 296 HPNLTIKLPARKIEEHKINDLIMKATQKQPVYLRPPYKATNATVKTIVTGLKEMLSVDT 355  
HPNLT L +I+ +++ TN I KA G+K YLRPPYKATNATV+ +GL +MLH+VDT

Sbjct: 303 HPNLTINLSVSEIQHVNMTQALEKACGKKRYLRPPYKATNATVQSSGLTQMLNTVDPT 362

-1603-

Query: 356 EDWKNHNTQAMMTNKKQLRPGGVLMHDIHQTTTDALETFIMDYLTGQYFVTVGELY 415  
 DW+NH-T +MTN+K QL+PGGV+LMHDIHQTTT+ALEFT+M+YL +GY VTV ELY+  
 Sbjct: 363 EDWENHSTGIMINVKQLQPGGVLMHDIHQTTTINALFTVMEYLKAEQYECVTVSELYA 422

- 5 GBS281d was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 8-10; MW 71.5kDa) and in Figure 187 (lane 10; MW 71kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 12; MW 46.5kDa) and in Figure 183 (lane 2; MW 46kDa). Purified GBS281d-GST is shown in lane 6 of Figure 237.
- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1453

A DNA sequence (GBSx1539) was identified in *S. agalactiae* <SEQ ID 4461> which encodes the amino acid sequence <SEQ ID 4462>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 57  
 >>> Seems to have no N-terminal signal sequence
- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2488 (Affirmative) < succ>  
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4463> which encodes the amino acid sequence <SEQ ID 4464>. Analysis of this protein sequence reveals the following:

- 25 Possible site: 40  
 >>> Seems to have no N-terminal signal sequence
- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2799 (Affirmative) < succ>  
 30 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 311/475 (65%), Positives = 389/475 (81%)
- 35 Query: 1 MTKEYQNTYNGEWKSSVNGEITLSPIDSSLGVPVPMTRREEVDHAKKAGREALFAWALT 60  
 + K+Y+N YKGEWK S N+I I +P LG VPMTR+ EVD + +AL W AL+  
 Sbjct: 1 LAQYKXNLVNGEKLSENEITTYAPATGEELGSPVPMTCQAEVDVAVASAKHLEDRALS 60
- 40 Query: 61 VYERQYLHKAADI+ IERDKKEIATVLAKEIKAYNASVTEVVRTADLIRYAAEGRILST 120  
 ERA YLHKAADI+ RD E+I +L+KE+K + A+V+EV+RTA+I YAAEGR+R+  
 Sbjct: 61 VYVRAAYLHKAADILVRDAEKIGAILSKVAKGHKAANSEVIRTARTINVAAREGLRWEG 120
- 45 Query: 121 SADGGHMDASTGUKLAVIRRPVGVILVLAIPYNPVNLGSKIAPALIGCVVMPKPT 180  
 EGG A++ K+A+RR+PVG+VLAIP+P+NYPVNL+GSKIAPALI GRVV KPPT  
 Sbjct: 121 EVLGGSGPFAASKKKLAIVRREPVGILVLAISPNNYPVNLGSKIAPALIGNVVALKPPT 180
- 50 Query: 181 QGSVSGILVLAFAEAGLPAAGVFNITIGRGSSEIGDYVHEEYVNFINTGSTPFGKRIK 240  
 QGS+SGIL+LA+AFARAG+PAGVFNITIGRGS IGDIYVHKH V+FINPTGSTP+G+ IGK  
 Sbjct: 181 QGSISGLLAEFAEAGIPAGVFNITIGRGSVIGDYVHEEYVNFINTGSTPIGBGIGK 240
- Query: 241 LAQMRPIMLELGGKIDAGVVLADADLNAAKQIVAGRYDYSQCRCTAIVKRVLVVEEVADEL 300  
 LAQMRPIMLELGGKD+ +VL DADL AAK IVAGA+ YSQCRCTA+KRVLV++VAD+L  
 Sbjct: 241 LAQMRPIMLELGGKDAIVLEADLALAAKQIVAGRYDYSQCRCTAIVKRVLVNDKVADEL 300
- 55 Query: 301 AEKISENVAKLVGDDPFDNATVTPVIDDMSADPTESLVVDARQKGAKEAFKPKDRGLRT 360

A +I V KLSVG P D+A +TP>ID ++ADF+E L+ DA KGA L F R+G L++  
 Subject: 301 AAEIKTLVEKLSVGMPEDDADITPLIDTSAADFVEGLIKDATDKGATALTAFNREGNLIS 360

Query: 361 PGLFDHVTITLDMKLAWEEPFGPILPIIRVKDAEEAVAANKSDPGLQSSVFTDRDFQKAFDI 420  
5 P LFDHVT DM+LAWEEKFGP+LPIIRV EEA+ I+N+S++GLQ+S+FT +F KAF I  
Sbjct: 361 PVLFDHVTITLMLAWEEPFGPVLPIIRVTTVEEAIKISNESEYGLQASIFTTNFPKAFGI 420

Query: 421 ANKLEVGTVHINNKTGRGPDNFPFLGLKSGSAGVQGI RYSIEAMTNVKSIVFDMK 475  
A +LEVGT VH+NNKT RG DNF PFLG K SGAGVQG++YSIEAMT VKS+VFD++  
10 Subject: 421 AEOLEVGTVHLLNNKTRGTDNFPFLGAKKSGSAGVQGVKYSIEAMITTVKSVVFDIQ 475

```

15 Lipop: Possible site: -1  Crend: 3
    Mcg: Discrim Score: -15.11
    GvK: Signal Score (-7.5): 0.17
        Possible site: 57
    >>> Seems to have no N-terminal signal sequence
    ALQM program count: 0 value: 1.22 threshold: 0.0
20 PERIPHERAL Likelihood = 1.22 187
    modified ALQM score: -0.74

*** Reasoning Step: 3

```

```

25      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

66.8/82.6% over 474aa

Streptococcus mutans  
EGAD|42413| NADP-dependent glyceraldehyde-3-phosphate dehydrogenase Insert characterized  
35 EGAD|42413|110509 NADP-dependent glyceraldehyde-3-phosphate dehydrogenase Insert  
characterized  
SP|SP9931|GAPN\_STIMU NADP-dependent GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (BC 1.2.1.19)  
NADP-DEHYDROPHOSPHATE GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (GLYCERALDEHYDE-3-PHOSPHATE  
DEHYDROGENASE [NADP+]) (TRIOSEPHOSPHATE DEHYDROGENASE) Edit characterized  
40 GP|642667|gb|AAA91091.1|138521 NADP-dependent glyceraldehyde-3-phosphate dehydro Insert  
characterized

ORF1989(3681 - 1725 of 2025)  
 RGD[42413|44796(1 - 475 of 475) NADP-dependent glyceraldhyde-3-phosphate dehydrogenase  
 (Streptococcus mutans)RGD[42413|110509 NADP-dependent glyceraldhyde-3-phosphate  
 dehydrogenase (Streptococcus mutans)SP[CG59931|GAPN\_STRMU NADP-DEPENDENT GLYCERALDHYDE-3-  
 PHOSPHATE DEHYDROGENASE (BC 1,2,1,9) (NON-PHOSPHORYLATING GLYCERALDHYDE-3-PHOSPHATE  
 DEHYDROGENASE) (GLYCERALDHYDE-3-PHOSPHATE DEHYDROGENASE [NADP+]) (TRIOSEPHOSPHATE  
 DEHYDROGENASE).GP[642667|gb|AAA91091.1||L38521 NADP-dependent glyceraldhyde-3-phosphate  
 Watch = 49.3  
 %Identity = 66.7 %Similarity = 82.5  
 Matches = 317 Mismatches = 93 Conservative Sub.s = 75

[illegible]

65



-1606-

Query: 121 DNFYMQERAADIRDAKRVLAHLIGVKLPNPATINEESIVIAHDLTPSDTAQLNQFVKA 180  
 DNFYMQERAADIRDAKRVLAHLIG KLPNPATI+EEISVIAHDLTPSDTAQLNQFVKA  
 Sbjct: 121 DNFYMQERAADIRDAKRVLAHLIGAKLPNPATIDEESIVIAHDLTPSDTAQLNQFVKA 180

5 Query: 181 FVTNIGGRTSHSAIMARTLEIAAVALGTNDITERVQDGLIAVNGITGEVIIIPTEAQISA 240  
 FVTNIGGRTSHSAIMARTLEIAAVALGTNDIT RV+DG ++AVNGITGEVII PT+ Q++  
 Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVALGTNDITSRVKDGIVAVNGITGEVIIINPTDQVAE 240

10 Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFEAANIGTPKDVGVNMGAEAVGLYRTE 300  
 FKAAGEAYAKQKAEWALLKDA+TVTADGKHFEAANIGTPKDVGVNMGAEAVGLYRTE  
 Sbjct: 241 FKAAGEAYAKQKAEWALLKDAKTVTADGKHFEAANIGTPKDVGVNMGAEAVGLYRTE 300

15 Query: 301 FLYMDSQDFPTEDEQYAYKAVLBGMNGKPVVVRTMDIGDGKELPYDLPKEMNPLFGR 360  
 FLYMDSQDFPTEDEQYAYKAVLBGMNGKPVVVRTMDIGDGKELPY DLPKEMNPLFGR  
 Sbjct: 301 FLYMDSQDFPTEDEQYAYKAVLBGMNGKPVVVRTMDIGDGKELPYDLPKEMNPLFGR 360

20 Query: 361 ALRISISSETGDMFRQTRALLRASVHGQLRIMFPVALLKEFRAAKAIFPEEKAMLLAD 420  
 ALRISISSETG+AMFRQTRALLRASVHGQLRIMFPVALLKEFRAAKAIF+EEKAM L A+  
 Sbjct: 361 ALRISISSETGNAMFRQTRALLRASVHGQLRIMFPVALLKEFRAAKAIFDEEKAMLAE 420

25 Query: 421 GVAVAGIEVGIWIRIPAAAMLADQFAKEVDFFSISGTNDLIQYTMAADRNBNQVSYLYQP 480  
 GVAV++ I+VGIMIRIPAAAMLADQFAKEVDFFSISGTNDLIQYTMAADRNBNQVSYLYQP  
 Sbjct: 421 GVAVSDIQVGIWIRIPAAAMLADQFAKEVDFFSISGTNDLIQYTMAADRNBNQVSYLYQP 480

30 Query: 481 YNFSILRLDNNVIAKAHAGKWMCMCEMAGDQTAAPLVLMGMLDEFSMSATSVLRTSL 540  
 YNFSILRLDNNVIAKAHAGKWM CMCEMAGDQ AVPLLV MGLDEFSMSATSVLRTSL  
 Sbjct: 481 YNFSILRLDNNVIAKAHAGKWMCMCEMAGDQKAVPLLVLMGMLDEFSMSATSVLRTSL 540

30 Query: 541 MKKLDTARMQEVANRAL+ECSTMEEV-EL KEV+ D 577  
 MKKLDTARM+EVANRAL+ECSTMEEV-EL KEV+ D  
 Sbjct: 541 MKKLDTARMQEVANRALTECSTMEEVLELSKEYVND 577

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4467> which encodes the amino acid  
 35 sequence <SEQ ID 4468>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.0875 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 540/577 (93%), Positives = 561/577 (96%)

Query: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTNAREARLDVALQASQDELGVIRE 60  
 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTV DTNAREARLDVALQ+QDELGVIRE  
 Sbjct: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVADTNAREARLDVALQAQDELGVIRE 60

50 Query: 61 NAVESLGEBAAAVFDAILMVLSDPEMINQIKETIRAKOVNARTGLKEVTDMFTTIFEGME 120  
 AVESLGEBAAAVFDAILMVL+DPEMI+Q+KETIRAKQ NAETGLKEVTDMFTTIFEGME  
 Sbjct: 61 NAVESLGEBAAAVFDAILMVLADPEMISQVSKETIRAKQNAETGLKEVTDMFTTIFEGME 120

55 Query: 121 DNFYMQERAADIRDAKRVLAHLIGVKLPNPATINEESIVIAHDLTPSDTAQLNQFVKA 180  
 DNFYMQERAADIRDAKRVLAHLIGVKLPNPATINEESIVIAHDLTPSDTAQLNQFVKA  
 Sbjct: 121 DNFYMQERAADIRDAKRVLAHLIGVKLPNPATINEESIVIAHDLTPSDTAQLNQFVKA 180

60 Query: 181 FVTNIGGRTSHSAIMARTLEIAAVALGTNDITERVQDGLIAVNGITGEVIIIPTEAQISA 240  
 FVTNIGGRTSHSAIMARTLEIAAVALGTNDIT+RV+DG +IAVNGITGEVII+P+E Q+ A  
 Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVALGTNDITRVKDGDIIVAVNGITGEVIIIPDSDQVLA 240

65 Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFEAANIGTPKDVGVNMGAEAVGLYRTE 300  
 FK AG AYAKQKAEM+L+KDA T TADGKHFEAANIGTPKDVGVNMGAEAVGLYRTE  
 Sbjct: 241 FKAAGEAYAKQKAEWALLKDAHTTADGKHFEAANIGTPKDVGVNMGAEAVGLYRTE 300



-1607-

```

Query: 301 FLYMDSQDFFTTSDQYEAYKAVLGGMNGKPVVVRTMDIGGKELPYFDLPKEMNPFILGPR 360
      FLYMDSQDFFTTSDQYEAYKAVLGGMNGKPVVVRTMDIGGKELPYFDLPKEMNPFILGPR
Sbjct: 301 FLYMDSQDFFTTSDQYEAYKAVLGGMNGKPVVVRTMDIGGKELPYFDLPKEMNPFILGPR 360

5   Query: 361 ALRISISSETGDMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEKAMILLAD 420
      ALRISISSETGDMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEKAMILLAD
Sbjct: 361 ALRISISSETGDMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEKAMILLAS 420

10  Query: 421 GVAVABGIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRNNQVSYSLYQP 480
      GVAVA+ 1+VGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRNNQVSYSLYQP
Sbjct: 421 GVAVADDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRNNQVSYSLYQP 480

15  Query: 481 YNPSILRLINNVIKAAHABGKWAGMCGEMAGDQAVPLLVGMGLDEFSMSATSVLRTRSL 540
      YNPSILRLINNVIKAAHABGKWAGMCGEMAGDQ AVPLLVGMGLDEFSMSATSVLRTRSL
Sbjct: 481 YNPSILRLINNVIKAAHABGKWAGMCGEMAGDQAVPLLVGMGLDEFSMSATSVLRTRSL 540

Query: 541 MKKLDATAKMEEYANRALSECSTMSEVIELQKEYVDFD 577
      MKLD+AKMEEYANRAL+ECST SEV+EL KEYV D
20  Sbjct: 541 MKKLDAAKMEEYANRALTECSTAEEVLELSKEYVSED 577

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1455

25 A DNA sequence (GBSx1541) was identified in *Sagalactiae* <SEQ ID 4469> which encodes the amino acid sequence <SEQ ID 4470>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

30  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35  The protein is similar to a protein from S.bovis:

>GP:BAA78048 GB:AB027569 histidine containing protein [Streptococcus bovis]
Identities = 86/87 (98%), Positives = 87/87 (99%)

40  Query: 1 MASQDFHIVARTGIHARPATLLVQTASKFASDITLDYKGAVALNKSIMGVMSLGVGQGD 60
      MASQDFHIVARTGIHARPATLLVQTASKFASDITLDYKGAVALNKSIMGVMSLGVGQGD
Sbjct: 1 MASQDFHIVARTGIHARPATLLVQTASKFASDITLDYKGAVALNKSIMGVMSLGVGQGD 60

Query: 61 VTISABGADADDAIAAIEETMTKEGLA 87
      VTISABGADADDAIAAIEETMTKEGLA
45  Sbjct: 61 VTISABGADADDAIAAIEETMTKEGLA 87

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4471> which encodes the amino acid sequence <SEQ ID 4472>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

50  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/87 (98%), Positives = 87/87 (99%)

-1608-

Query: 1 MASKDPHIVARTGIHARDPILLVQTASKFASDITLDYKGAVNLEKINGVMSLGVQGGAD 60  
 MASKDPHIVARTGIHARDPILLVQTASKFASDITLDYKGAVNLEKINGVMSLGVQGGAD  
 Sbjct: 1 MASKDPHIVARTGIHARDPILLVQTASKFASDITLDYKGAVNLEKINGVMSLGVQGGAD 60

5 Query: 61 VTISAEGADADDAIAAIEETMIKKGIA 87  
 VTISAEGADR-DAIAAIEETMIKKGIA  
 Sbjct: 61 VTISAEGADADDAIAAIEETMIKKGIA 87

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 10 vaccines or diagnostics.

**Example 1456**

A DNA sequence (GBSx1542) was identified in *S. agalactiae* <SEQ ID 4473> which encodes the amino  
 acid sequence <SEQ ID 4474>. This protein is predicted to be glutaredoxin-like protein nrhd (b2673).  
 Analysis of this protein sequence reveals the following:

15 Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.4532 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAA63372 GB:X92690 glutaredoxin-like protein [Lactococcus  
 lactis]  
 Identities = 42/70 (60%), Positives = 53/70 (75%)

Query: 4 ITTVFSKNNCMQCKMTKKFLDQHGADPEEINIDEKPEKIEYVKNLGPSAAPVIEAGNVVFS 63  
 +TV+SKNNCMQCKM KK+L +H F EINIDE+EE +E V +GP AAFVI + PS  
 30 Sbjct: 2 ITTVFSKNNCMQCKMTKKFLDQHGADPEEINIDEKPEFVEKVIEMGPRAAFVITKDDFAFS 61

Query: 64 GFQPSKLKEL 73  
 GF+PS+L +L  
 35 Sbjct: 62 GFPSSELAKL 71

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4475> which encodes the amino acid  
 sequence <SEQ ID 4476>. Analysis of this protein sequence reveals the following:

40 Possible site: 17  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.4606 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 56/71 (78%), Positives = 68/71 (94%)

50 Query: 4 ITTVFSKNNCMQCKMTKKFLDQHGADPEEINIDEKPEKIEYVKNLGPSAAPVIEAGNVVFS 63  
 IT+SKNNCMQCKMTKKFL+HG +P+EINIDE PEK++YVK+LGP+AAFPVIEA N+VFS  
 Sbjct: 13 ITTVFSKNNCMQCKMTKKFLDQGVNPGSEINIDHPEKVDYVKSGLGTSAPVIEADNLVFS 72

Query: 64 GFQPSKLKELV 74  
 GFQP+KIKKI+  
 55 Sbjct: 73 GFQPAKLKELI 83

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1457

A DNA sequence (GBSx1543) was identified in *S. agalactiae* <SEQ ID 4477> which encodes the amino acid sequence <SEQ ID 4478>. This protein is predicted to be ribonucleotide reductase subunit R1E (nrde). Analysis of this protein sequence reveals the following:

Possible site: 49  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3676 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD41036 GB:AF112535 ribonucleotide reductase alpha-chain  
[Corynebacterium glutamicum]  
Identities = 366/701 (52%), Positives = 488/701 (69%), Gaps = 19/701 (2%)

Query: 23 NGQIPHLKDKREALTAFFKENVQPNKAFDSITDKIAYLLKYDYLEEAFINKYRPEFIEEL 82  
NG+I KD+EA +F ++V N+ F ++ +KI YL++ Y + L+KY +FI++L  
Sbjct: 22 NGKIQFEKDEAANQYFLQHVNTVFHNLQEKIDYLVENKYDIPVLVDKYPQFIKDL 81

Query: 83 STKLFDKKFRFKSFMAYKFYQYALKTNDGEYLLSIEDRVLFNALYPADGDEELATDL 142  
+ + KFRF+SF+ AYK+Y Y LKT DG YLE EDRV AL ADGD LA +L  
Sbjct: 82 FKRAYGPKFRFQSGFLGAYKTYTTLKTDFGRYLERFEDRVCMVALTLADGDRALAEHL 141

Query: 143 ALEMISQRYQATPSPFINAGRSRRGELVSCFLIQVIDMNAIGRSINSALQLSRIQGGVG 202  
E++S R+QFATP+FLN+G+++RGE VSCFL+++ D+M +IGRSINSALQLS+ GGGV  
Sbjct: 142 VDEIMSRGFQATPTPFLNKGAKRGEGFVSCFLRIEDNMESIGRSINSALQLSKRGGVA 201

Query: 203 ISLSNLRAGAPINGPAGAAAGVVPVVKLPEDSFYSNQLQRRQAGGVYLDVHPDIIS 262  
+ LSNLRAGAPIK +SGV+PVMKL ED+FSY+NQLG RQAG VYL+ HPDI+S  
Sbjct: 202 LLLSNLRAGAPIKKIENQSGGVIPVMKLEDAFSYANQLGARQAGAVYINAHHPDIIS 261

Query: 263 FLSTKKENADEKVRVKTLISGITVPDKFYELARNQENYLFSPYSISIEREYGVPPSYIDIT 322  
FL TK+ENADEK+R+KTLISLG+ +PD +ELA+ N+MYLFSY +ER YG PF+ + IT  
Sbjct: 262 FLDTKRENADEKIRIKTISLGVVIPDITFRIAKRNDMYLFSYDVERIYQKPFADVSIT 321

Query: 232 EKYDELVANPNITKTKINARDLETRISKLQESGYPIYINIDTANRTNPVDGKIIMSNLC 382  
E YDE+V + I KTKINAR ++++Q ESGYPYI+ DT N +NP++G+I SNLC  
Sbjct: 322 EHYDEMVDDDRIRKTKINARQFPQTLAEIQESGYPIYIMYRDTVNASNPIEGRITHSNLC 381

Query: 383 SEILQVQKPSLINDAEOYLEMGTDISCLNSTVNINMMTSPDGKSIKIMTRALTFVTD 442  
SEILQV PS ND Y E+G DISCHLS NV M SP+F K+I+T R LT V++  
Sbjct: 382 SEILQVSTPSRFNDDLYAEVGEDISCHLSNVAMMDSNPFKTIETATROLTVAESEQ 441

Query: 443 SNIRAVPTIKNGNAQAHTFGLGAMGLHSYLAQNHIEYGSPEISLEFDTIYFMAQNYLWLE 502  
++I++VP+I+ GN AH GLG M LH Y + H+ YGS E++FT+ YF + Y L  
Sbjct: 442 TSIDSVPFIRKNGEAAHAGLQGMNHLGYFGRHHMYGSEALDFTINAYFAVLYQCLSA 501

Query: 503 SNNIARERQTFVFGPEKSYADGTYFDKYVSGKVPQSDKVKSLFA--NHFIPEAKDWN 560  
SN IA ER F FE SKYA G YPD + + F P+SDKVK LFA N P +DW  
Sbjct: 502 SNKIATERGERGFKNFENSKYA+GGEYDPPDNDANFAPKSDKVKELFAKSNIHPTVDEWNA 561

Query: 561 LRYAVMKDGLYHQNRLAVAPNGSISYINDCSASHPITQRIEERQEKIKITTYYPANGLA 620  
L+ VM+ GL++N AV P GSISYIN+ ++SHPI +IE R+E KIG++YYP+ +  
Sbjct: 562 LKADVMHGLFPNNLQAVEPPTGSISYINNSTSSHPITASKIEIRKBEKIGRVYYPAPRND 621

Query: 621 TDTPIPYTTSAYMDMKRVLDVYAAATEHVDQGLSMFLRSELPKELYBWRTEKQOTD 680  
D + Y+ AY++ K+ID YA AT++VDQGLS+TLF + TTRD  
Sbjct: 622 NDNLEYFEDAYBIGYKIIDTYAVATKYVDQGLSLTLFFK-----DTATTD 668

-1610-

Query: 681 LSLILNRYAPNKGVSIIYYI--RTPTDDGSEVGANQCESCVI 719  
 ++ + YA+ KG+K+YII R +G+EV + C SC++  
 Sbjct: 669 INKAIQIYAMRGIKTLYIIRLQVALSGTGV--DGCVSQML 707

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4479> which encodes the amino acid sequence <SEQ ID 4480>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 628/719 (87%), Positives = 682/719 (94%)

Query: 1 MSLLQIGDVSYPRINNEINRPVNGQIPLEKDKRALTAFFKENVQPNKAFDSITDKIAYL 60  
 MSLK++GD+SYFRINNEINRPVNG+IPLEKDKRAL AF ENV FN+ +F SIT+KI YL  
 Sbjct: 1 MSLLKIDISISYFRINNEINRPVNGKIPLEKDKRALKAFSAENVLPMTMSPTSITKEIYL 60

Query: 61 LKYDYLESAFLNKYRPEFISELSTKLFDKKFKRPSMAAYKFYQQYALKTNIDGEYLLSI 120  
 + DY+E AF+ KYRPEFI EL + + + FAFKSPMAAYKFYQQYALKTNIDGE+YLE++  
 Sbjct: 61 ISNDYIESAFIQYRPEFITEIDSIITKSENFAPKSPMAAYKFYQQYALKTNIDGSHYLEN 120

Query: 121 EDRVLFNALYFADGDEKLATDLALEMISRYQPATPSFLNAGRSRRGELVSCFLIQVTDD 180  
 EDRVLFNALYFADG E+LA DLA+EMI+QRYQATPSFLNAGRSRRGELVSCFLIQVTDD  
 Sbjct: 121 EDRVLFNALYFADGDEKLADLALEMINKRYQATPSFLNAGRSRRGELVSCFLIQVTDD 180

Query: 181 MNAIGRSINSALQLSRIGGGVGLSLNLEAGAPIKGFAGAASGVVPMKLFDSFSYSN 240  
 MN+IGRSINSALQLSRIGGGVGL+LSNLEAGAPIKG+AGAASGVVPMKLFDSFSYSN  
 Sbjct: 181 MNSIGRSINSALQLSRIGGGVGLTSLNLEAGAPIKGYAGAASGVVPMKLFDSFSYSN 240

Query: 241 QLQQRQAGGVYLDVFHPDIIIFLSTKKNADKRVKLTSLGITVPOKPYELAR N++M 300  
 QLQQRQAGGVYLD+VFHPDII+FLSTKKNADKRVKLTSLGITVPOKPYELAR N++M  
 Sbjct: 241 QLQQRQAGGVYLDVLFHPDIIIFLSTKKNADKRVKLTSLGITVPOKPYELARNGEDM 300

Query: 301 YLFSYPSIEREYGVFFSYIDITKYLDELVANPMTTKINARDLETSIKLQCESGYPIY 360  
 YLFSY++E+EYGF+FF+Y+DIT YDELVANP IYTKKI ARDLETSIKLQCESGYPIY  
 Sbjct: 301 YLFSYINVEYGIFFNYLDTINMDELVANPKITTKIKARDLETSIKLQCESGYPIY 360

Query: 361 INIDTANRINFPVQKIIIMSNLSEILQVQKPSLINDAGEYLEMGTDISCNLQSTVLNRM 420  
 INIDTAN+ NF+DGKIIIMSNLSEILQVQ PSINDAGE++EMGTDISCNLQSTN+LNRM  
 Sbjct: 361 INIDTANKANFDGKIIIMSNLSEILQVQVPSLINDAGEFVEMGTDISCNLQSTN+LNRM 420

Query: 421 TSPDFGSKITMTRALTFTVTDSSIEAVPTIKHNGSQAHITGLGAMGLHSYLAQHIEYG 480  
 TSPDFG+SKITMTRALTFTVTDSS+IRAVPTIK+GN+QAHITGLGAMGLHSYLA++HIEYG  
 Sbjct: 421 TSPDFGSKITMTRALTFTVTDSSIEAVPTIKHNGSQAHITGLGAMGLHSYLAQHIEYG 480

Query: 481 SPSSIEPTDIYFMLNWIWTLVESNNIAREKQITTVGFENSKYADGTYPKCVSGHFPQS 540  
 SPSSIEPTDIYFML+NWITLVESNNIAREKQITTVGFEN SKIA+G+YDKTV+G FVP+S  
 Sbjct: 481 SPSSIEPTDIYFMLNWIWTLVESNNIAREKQITTVGFENSKYANGSYFDKTVGHFVPS 540

Query: 541 DKVKSLEPNHPIPEAKIMENLRYAVMKDGLYHQRLAVAPNGSISYINDCSASIHPTQR 600  
 D V K LF +HPIP+A DWE LR AV KDGLYHQRLAVAPNGSISYINDCSASIHPTQR  
 Sbjct: 541 DLVKDLFKHPIPGASDWEALRDVQKDGLYHQRLAVAPNGSISYINDCSASIHPTQR 600

Query: 601 IEERQEKIKGIYYPANGLATDTIPYPTTSAYDMMKRVIDVYAATHEVDQGLSLTLFLR 660  
 IEERQEKIKGIYYPANGL+TDTIPYPTTSAYDMMKRVIDVYAATHEVDQGLS+TLTLR  
 Sbjct: 601 IEERQEKIKGIYYPANGLSTDTIPYPTTSAYDMMKRVIDVYAATHEVDQGLSLTLFLR 660

Query: 661 SELPKHLYEWKTSKQTTDRDLSILRNYPNKGVSIIYYIRTPPTDDGSEVGANQCESCVI 719  
 SELP ELVWKT+SKQTTDRDLSILRNYPNKG+KSIIYYIRTPPTDDG EVGANQCESCVI  
 Sbjct: 661 SELPKHLYEWKTSKQTTDRDLSILRNYPNKGISIIYYIRTPPTDDGSEVGANQCESCVI 719

-1611-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1458

- 5 A DNA sequence (GBSx1544) was identified in *S. agalactiae* <SEQ ID 4481> which encodes the amino acid sequence <SEQ ID 4482>. This protein is predicted to be ribonucleotide reductase subunit R2F (nrdb). Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4583 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

15 A related GBS nucleic acid sequence <SEQ ID 9753> which encodes amino acid sequence <SEQ ID 9754> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:RA014561 GB:AF050168 ribonucleoside diphosphate reductase small
subunit [Corynebacterium ammoniagenes]
Identities = 166/313 (53%), Positives = 215/313 (68%), Gaps = 1/313 (0%)

Query: 10 EAINNNEIEDVIDKSTWEKLTQFQWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFQGLTLL 69
+AINNN I D D W++LT FWL +IP+SND+ W K++ QE+ +VF GLTLL
Sbjct: 17 KAINNNVLPDEKLEVDRLAIGNFWLPEKIPVSNDIQSNWMTPOBQLATRVPTGLTLL 76

Query: 70 DTMOSETGVEAIRADVPTPHEEAVLANNIQFMESVHAKSYSSIFSTIANTKSEIEEIEFWTN 129
DT+Q G ++ DV T HEE V NI FMESVHAKSY+IF TL + +I E F W+
Sbjct: 77 DTIQGTGVAISLLPDVETMHEEGVYTNIAFMESVHAKSYSNIFMTLASTPQINEAFWSE 136

Query: 130 NNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFTPLLYGNKNNLANVAEIIK 189
NE LQ KA+II Y + L+KKVAST LE+FLFYSGF+ P+Y KL N A+II+
Sbjct: 137 ENENLQKAKIIMSYNGDDPLKQKVASTLLESFLFYSGFTPLMYLSSRAKLTNTADIIR 196

Query: 190 LIIRDESVMGTIYGKFKQLGPNELPRDEQENFRDWMYDLLQLYENEYKTKLYDGVGW 249
LIIRDESVMG YIGYK+Q G +L E EQE ++ + +DL+Y LYENE +YT+ +YD +GW
Sbjct: 197 LIIRDESVMGTIYGKFKQGVKLSBAQBEYKAYTFDMLYDLYENEIEYTEDITDVLGW 256

Query: 250 TEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPVWIGIS-TGTSNHDFFPSQVNGYLL 308
TE+V FLRYNANKAL NIG + LFP V+P +++ +S NHDFFS G+ Y++
Sbjct: 257 TEDVKRFLRYNANKALMNLGYGLFPPTDETQVSPAILSSLSNADENHDFFPSGSGSYVI 316

Query: 309 GSVEAMHDDDDYNY 321
G E DDD+++
Sbjct: 317 GKAEOTDDDDND 329

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4483> which encodes the amino acid sequence <SEQ ID 4484>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4583 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-1612-

Identities = 315/319 (98%), Positives = 316/319 (98%)

```

Query: 5  MTTYYSAINWNEIEDVIDKSTWEKLTQFWLDTIRFLSNDLDDWRKLSQEKDLVGKVPF 64
5  Sbjct: 1  MTTYYSAINWNEIEDVIDKSTWEKLTQFWLDTIRFLSNDLDDWRKLSQEKDLVGKVPF 60

Query: 65  GLTLLDTMQSETGVBAIRADVKTPEKRAVLNNIQPMBSVHAKSYSSIPSTLATKKEIEEI 124
10 Sbjct: 61 GLTLLDTMQSETGVBAIRADVKTPEKRAVLNNIQPMBSVHAKSYSSIPSTLATKKEIEEI 120

Query: 125  FEWTNNNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFPTFLYYLGNKKLANV 184
15 Sbjct: 121 FEWTNNNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFPTFLYYLGNKKLANV 180

Query: 185  ARIIKLIIRDSEVHGTYIGYKFLGPNELFEDQENFRDMYDLLYQLYENEEKYTKILY 244
20 Sbjct: 181 ARIIKLIIRDSEVHGTYIGYKFLGPNELFEDQENFRDMYDLLYQLYENEEKYTKILY 240

Query: 245  DGVGWTEEVMTFLRYNANKALMNLGQDLFFDTANDVNFVWNGISTGTSHNDDFFSQVGN 304
25 Sbjct: 241 DGVGWTEEVMTFLRYNANKALMNLGQDLFFDTANDVNFVWNGISTGTSHNDDFFSQVGN 300

Query: 305  GYLLGSVEAMHDDYNYGL 323
30 Sbjct: 301 GYLLGSVEAMHDDYNYGL 319

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1459

30 A DNA sequence (GBSx1545) was identified in *S. agalactiae* <SEQ ID 4485> which encodes the amino acid sequence <SEQ ID 4486>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence
35  INTEGRAL    Likelihood = -0.27    Transmembrane    50 - 66 ( 50 - 66)

----- Final Results -----
          bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 45 Example 1460

A DNA sequence (GBSx1546) was identified in *S. agalactiae* <SEQ ID 4487> which encodes the amino acid sequence <SEQ ID 4488>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence
50  INTEGRAL    Likelihood =-14.38    Transmembrane    176 - 192 ( 168 - 201)
          Likelihood = -4.57    Transmembrane    25 - 41 ( 22 - 42)
          Likelihood = -3.88    Transmembrane    94 - 110 ( 94 - 112)
          Likelihood = -1.49    Transmembrane    70 - 86 ( 70 - 86)
          Likelihood = -1.01    Transmembrane    128 - 144 ( 128 - 144)
55  ----- Final Results -----

```

-1613-

bacterial membrane --- Certainty=0.6753 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9751> which encodes amino acid sequence <SEQ ID 9752> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 55/184 (29%), Positives = 98/184 (52%), Gaps = 4/184 (2%)
- 10 Query: 16 MSKNNNTTCLIEITAFALAMALSMIP----DFASWPTSPGAIFLILFAIRRGTKYGLF 71  
 M+++ LIE AI A A+ L ++ + S IP+ L + R G K GL  
 Sbjct: 1 MNQSKQLVRLIEIAIMTAAAVILDIVSGMFLSMPOGGSVSIWMIPIFLISPRNGVKAGLT 60
- 15 Query: 72 AGLTWGLMHFVLSKVYYLSQVFIEYILAFISMGLAGVFSAPKFDALSSSSKTKALSIA 131  
 GL+ GL+ + ++ Q+ ++YI+AF ++G+G F++ + A S +K K +  
 Sbjct: 61 TGLTGLVQIAIGNLFAQHFQVQLLLDYIVAFAAIGISGCFASSVRKAASVSTKGLIVSV 120
- 20 Query: 132 LSGAILATLVRYVMFYIAGVIFWASYPKMSATLSVLSVNGTAGLLTLFPVVISIIILV 191  
 +S + +L+RY H I+G +P+ S+APKG +YSL+ N T + + I + +L  
 Sbjct: 121 VSAVFIGSLRLRYAAHVISGAVFFGSPAPKGTFPWVIYSLTYNATYMPVSPITCAIVLCILF 180
- Query: 192 ISYP 195  
 ++ P  
 25 Sbjct: 181 MTAP 184

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4489> which encodes the amino acid sequence <SEQ ID 4490>. Analysis of this protein sequence reveals the following:

- Possible site: 20  
 >>> Seems to have a cleavable N-term signal seq.
- |          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -9.34 | Transmembrane | 162 - 178 ( 156 - 183) |
| INTEGRAL | Likelihood = -9.34 | Transmembrane | 110 - 126 ( 107 - 130) |
| INTEGRAL | Likelihood = -1.22 | Transmembrane | 55 - 71 ( 55 - 71)     |
- 35 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4736 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- 40 The protein has homology with the following sequences in the databases:
- >GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 55/189 (29%), Positives = 100/189 (52%), Gaps = 10/189 (5%)
- 45 Query: 1 MSFNINVKYLIEAAIFALAMTLSPIDFAGNF--SPSYGATLAV----IPSIIRGLKY 53  
 M+ + + LIE AI A A+ L + +G P P G++++ + S R G+K  
 Sbjct: 1 MNQSKQLVRLIEIAIMTAAAVILDIV---SGMFLSMPOGGSVINMIPFLISPRNGVKA 57
- Query: 54 GMLAGLIGNLHFLVLFQKVYLSMSQVFIEYILAFISMGLAGVSPDSLIKTLRQQTFFAV 113  
 G+ GL+ GL+ +G ++ Q+ ++YI+AF ++G+G F+ S+ K + +  
 50 Sbjct: 58 GLITGLTGLVQIAIGNLFAQHFQVQLLLDYIVAFAAIGISGCFASSVRKAASVSTKGLII 117
- Query: 114 FLAIMASLAVTVRYLMHFLAGIIFWGSYAPKMSAVWYFVSVMGTAGVLTFLITCLAIM 173  
 + A + +RY H ++G +F+GS+APKG YS + N T V +F+I + L  
 Sbjct: 118 VSVSAVFIGSLRLRYAAHVISGAVFFGSPAPKGTFPWVIYSLTYNATYMPVSPITCAIVLC 177
- 55 Query: 174 IALPIHPQL 182  
 + P+L  
 Sbjct: 178 LLFWIAPRL 186

- 60 An alignment of the GAS and GBS proteins is shown below.

Identities = 116/186 (62%), Positives = 138/186 (73%)

-1614-

Query: 16 MSKNNNTTCLIEITAIPAALAMALSMIPDFASWFTSPGAIPILILFALRGTKYKGLFAGLI 75  
 MS N N LIE AIPAALAM LS IPDFA WF+PS+GAI L++P+LRRG KYG+ AGLI  
 5 Sbjct: 1 MSFNTNWKYLIEAATIPAALAMTSLSPIDFAGWFSPSYGAIALVIFSLRRGLKYKGLAGLI 60

Query: 76 WGLLHFVLSKVYYLSLSQVFIETILAFISMGLAGVPSAKFKDALSSSSKTKALSIALSGA 135  
 WGLLHFVL KVVYLS+SQVFIETILAF SMGLAG PS L A+ LA+ +  
 10 Sbjct: 61 WGLLHFV/LGKVVYLSMSQVFIETILAFISMGLAGSPSDSLIKTLERQQTFFAVFLAIMAS 120

Query: 136 ILATLVRYVWHYIAGVIFWASYPKGSATLYSLSVNGTAGLTLFFVVISIIILVISYP 195  
 +LA VRY+WH++AG+IFW SYAPKGSMA YS SVNGTAG+LT ++++I + +P  
 10 Sbjct: 121 LIAVTVRYLMHFLAGII FWGSYAPKGSMAVWYSFVNGTAGVLTFLITCLALMALPIHP 180

Query: 196 SFFLEK 201  
 F PK  
 15 Sbjct: 161 QLFDEK 186

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 20 Example 1461

A DNA sequence (GBSx1547) was identified in *S. agalactiae* <SEQ ID 4491> which encodes the amino acid sequence <SEQ ID 4492>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

25	INTEGRAL	Likelihood = -7.43	Transmembrane	206 - 222 ( 199 - 223)
	INTEGRAL	Likelihood = -6.64	Transmembrane	24 - 40 ( 19 - 42)
	INTEGRAL	Likelihood = -6.58	Transmembrane	61 - 77 ( 51 - 78)
	INTEGRAL	Likelihood = -6.58	Transmembrane	134 - 150 ( 132 - 154)
30	INTEGRAL	Likelihood = -4.62	Transmembrane	226 - 242 ( 224 - 245)
	INTEGRAL	Likelihood = -3.72	Transmembrane	107 - 123 ( 106 - 125)

----- Final Results -----  
 bacterial membrane --- Certainty=0.3972(Affirmative) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9749> which encodes amino acid sequence <SEQ ID 9750> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

40 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4493> which encodes the amino acid sequence <SEQ ID 4494>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

45	INTEGRAL	Likelihood = -10.46	Transmembrane	134 - 150 ( 131 - 159)
	INTEGRAL	Likelihood = -7.59	Transmembrane	107 - 123 ( 103 - 128)
	INTEGRAL	Likelihood = -7.48	Transmembrane	225 - 241 ( 213 - 248)
	INTEGRAL	Likelihood = -7.22	Transmembrane	205 - 221 ( 199 - 224)
50	INTEGRAL	Likelihood = -3.56	Transmembrane	50 - 66 ( 50 - 73)
	INTEGRAL	Likelihood = -1.28	Transmembrane	16 - 32 ( 16 - 33)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5182(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.



-1615-

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/253 (32%), Positives = 149/253 (58%), Gaps = 5/253 (1%)

```

5  Query: 6  IKQSDTTVPVRIKSLILGGFGAILGSGVGLFIIF--GQDKYLSRI--NIVQYPLMVSRI 61
      +K+ +F+R++K L+ G I+G + F+ G+ +L+ + + + + + + + + + + + + + + + + + + +
  Sbjct: 1  MKKKKNSPLRLKLLSSLAGGILGGMVGAFIYGHGRLDHLTFLKDDVNLILILNL 60

10 Query: 62  VVIITALPSLIYLYQIQKYQKVFVWDEBQ--SSEIYRQINLRHSYGMTFVVISISVLIVN 120
      VV+ S ++L Q++K V+ ++E SE YRQ+N +H+Y M +++ +LS+ N
  Sbjct: 61  VVVITLTLSPVFLTQKKCTAVYNTIREDDISKNGYQLNKKHAYTMILLIIVASILSMCN 120

15 Query: 121  TLFNYKLNIFDDSVTLVPIPIYDLSELLGLHITYFLKVVYRNIRGIMTVAPTLKELKNN 180
      L L L IP+ D+ L+ +++ +K Y IRG + P LKELK+N
  Sbjct: 121  VLLGLTLTNDSSHAMLAIPILDLILLMLVIPPQALAMKRYNAIRGTIVPFPVNLKELKHN 180

20 Query: 181  VLQDRAELESNYKMCDFIVMNLGFIPTIYFVLFVFFISFVQKVEIVAIITTSIHIIYI 240
      ++ LDEAEL++ +K F+ V++L+G I F++Y +LFF+ +VE+ AI++ I +Y+
  Sbjct: 181  IMLDRAELQAYHKTSFESVLSLNGVILPSLYVILFFVFLPTQVELTALNLVLQLYL 240

25 Query: 241  LIKSLKAARHFYR 253
      L+KS R FYR
  Sbjct: 241  LVKSATMTRQFYR 253

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1462

A DNA sequence (GBSx1548) was identified in *S.agalactiae* <SEQ ID 4495> which encodes the amino acid sequence <SEQ ID 4496>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 57
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.5172(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1463

A DNA sequence (GBSx1549) was identified in *S.agalactiae* <SEQ ID 4497> which encodes the amino acid sequence <SEQ ID 4498>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 26
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2059(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AACT6650 GB:AE000440
    UDP-D-glucose: (galactosyl)lipopolysaccharide

```

-1616-

glucosyltransferase [Escherichia coli K12]  
 Identities = 70/256 (27%), Positives = 121/256 (46%), Gaps = 14/256 (5%)

5 Query: 1 MNLLPSIDMYVDHFKVMSLYSLVQTKMKLEIYVLQKT---LKKRHTELIQTONLEV 56  
 +N+ + +D Y+D V + S+V ++ L+ Y++  
 Sbjct: 28 INVAYGVDANILDGWGVSTISIVLNNRHINLDFYIADVNDGFFQKIAKAEQMOLRIT 87

10 Query: 57 GYHPIIVGTEVFAQAFTTDRYPDTIYVRLAHKFLPETLDRILYLDADMCLNDFSSLYD 116  
 Y + T+ P T + +Y+RL A + L TLDRA+LYLDAD+ +C D S L  
 Sbjct: 88 LYS---INTDKLQCLPCTQVWSRAMYFRLFAQLGLTLDRLLYLDADVCKGDI SGLH 144

15 Query: 117 MELSDQLYAPASHNWDGKFLDYVHKLRANVELESSYFNTGVLANRLPAIRKVVHQITL 176  
 + L A A+ D + + RL + EL YPM+GV+ ++L + L  
 Sbjct: 145 LGING---AVAANVKDVSVMQSKAVSRLESDPELLQYFMSGVVYLDLKKMADAKITKAL 201

20 Query: 177 DYIMQNRGLRLPDDILNGLYANLVKPTPDEIYNYDARYSLIYQLKSRNEMDLEWVINH 236  
 +M PDQD++N L + +P E Y+ Y++ +LK + + + +I  
 Sbjct: 202 SILMSKDNVYKIFDQVMVVLKGMFLERE---YNTIYTIKSELKDKTHQNYKILTE 258

25 Query: 237 -TVFLHFAGRDKPWIK 251  
 T+ +H+ G KPM K  
 Sbjct: 259 STILLHYTGRTKPMK 274

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1464

A DNA sequence (GBSx1550) was identified in *S.galactiae* <SEQ ID 4499> which encodes the amino acid sequence <SEQ ID 4500>. Analysis of this protein sequence reveals the following:

30 Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

---- Final Results ----

35 bacterial cytoplasm --- Certainty=0.1406 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1465

A DNA sequence (GBSx1551) was identified in *S.galactiae* <SEQ ID 4501> which encodes the amino acid sequence <SEQ ID 4502>. Analysis of this protein sequence reveals the following:

45 Possible site: 54  
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.72	Transmembrane	7 - 23 ( 1 - 28)
INTEGRAL	Likelihood = -4.30	Transmembrane	222 - 238 ( 216 - 238)
INTEGRAL	Likelihood = -3.66	Transmembrane	151 - 167 ( 140 - 170)
50 INTEGRAL	Likelihood = -3.50	Transmembrane	35 - 51 ( 34 - 58)
INTEGRAL	Likelihood = -3.35	Transmembrane	71 - 87 ( 69 - 88)
INTEGRAL	Likelihood = -3.29	Transmembrane	113 - 129 ( 113 - 132)
INTEGRAL	Likelihood = -2.81	Transmembrane	170 - 186 ( 168 - 190)
INTEGRAL	Likelihood = -2.71	Transmembrane	198 - 214 ( 197 - 217)

----- Final Results -----

```
bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07774 GB:AP001520 unknown conserved protein [Bacillus halodurans]  
Identities = 84/242 (34%), Positives = 147/242 (60%), Gaps = 16/242 (6%)

Query: 1 MVGLGTVINIVILIIVGGFVGLFLKNFLKESLQKSLMQANGVAVLFISISGVLEKMMLVK 60  
MV +GTV+N I++ +GL +KN + E ++ +LMQA+G+A++ + + KM L +  
Subject: 1 MVGLGTVMQRAIVRAIIGLVKN-IPERVETTLMOAIGLAVLIGV----KMGLOT 54

Query: 61 SHLISNHTNMIIITLALGTVLGELLSLDSYIDKFGNYLKQKTGSGNDIKFVEAPVTSTCT 120  
 LI +I +L +G V+GE+++L+ +D G +++ K G D AFTT+T  
 Subject: 55 OFLI-----VICSLVIGGVIGGEMINLEKRLDHLGRWTIESKVGKKDGSIAFAVTTTLI 108

Query: 121 VCIGAMAVVGSIQDGLAADHSILFAKGMGLDMIIIAIMTVSLGKGFALFSAIPVALLQGSILT 180  
+GAMAV+G++ G+ DHS+L K +LD + + T +LG G LFS+PV L QGS+  
Subject: 109 YVVVGAMAVLGALDSGLRGDHSVLLTKALLDGFLLAILEFTSLGIGVLFSAIPVVLYOGSIA 168

Query: 181 IVAF----FMGSLINPSSLDYLNLVGNMLIFCVGVNLLFLNLINIKVINMLPAIILAILWGS 236  
+ A ++ + L S + ++ G ++I +G+NLL +NI+V N+LP++++ + +  
Subject: 169 LFASQIDQVPTALMDSFITEESATGGVMIVAIGLNILNVVNRVANLLPSIVIVAVLV 228

Query: 237 FI 238  
F+  
Subject: 229 EV 230

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1466

A DNA sequence (GBSx1552) was identified in *S. galactiae* <SEQ ID 4503> which encodes the amino acid sequence <SEQ ID 4504>. This protein is predicted to be alanyl-tRNA synthetase (alaS). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -4.41    Transmembrane    805 - 821 ( 804 - 822)
```

----- Final Results -----

```
bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04986 GB:AP001511 alanyl-tRNA synthetase [Bacillus halodurans]  
Identities = 482/885 (54%), Positives = 618/885 (69%), Gaps = 27/885 (3%)

Query: 1 MKELSSAQIRQMWLD FWKSKCHSVEPSANLVPVNDPTLLWINSQVATLKKYFDG SVIPEN 60  
MK L+SAQ+RQM+LDF+K KGH VEPSA+LVP +DP+ILLWINSQVATLKKYFDG VIPEN  
Subject: 1 MKYLPSAQRQMFLDFKKRKHGHDVPSASLVPDDPSLILWINSQVATLKKYFDG SVIPEN 60

Query: 61 PRITNAQKSIRTNDIENVGKTARHHTFEMLGNPSIGDYFRDEAIEWGFELLTSPWFDF 120  
PRITNAQKSIRTNDIENVGKTARHHT FEMLGNPSIGDYF++EAIEW +E LTS +W F  
Sbjct: 61 PRITNAQKSIRTNDIENVGKTARHHTFEMLGNPSIGDYFKERAI EWAEFLTSEKWIGF 120

Query: 121 PKDKLYMTYYPPDDKDSYNRWIA-CGVEPSHLVPIEDNFWRIGAGPSGPDTEIFFDRGEDF 179

-1618-

K+KL +T +P+D ++Y+ W G+ ++ +E NPW+IG GPSPG+TRIF+DRG ++  
 Sbjct: 121 DKEKLSVTVHPHDEFRASYWKEKIGIPERRIIRLBGNFWDIGBPGSGNWTETPYDRGPEY 180  
 5 Query: 180 -----DPENIGLRLAARDINDRYEITWNVLSQPNADPAVRSEYKELPNKNDIDVAGL 234  
 DFE L ENDRY+V+WN+V SQFN +P Y LP KNIDIG GL  
 Sbjct: 181 GDQPNDFE-----LYPGENDRYLEVWNLVFSQPNHFD---GSYTPLPKKNIDIGMEL 231  
 10 Query: 235 ERLAAV+MQGAKTINFETDLFWPIIREVEKLSGKTYDPDGD-NMSPKVIADHIALSFAIGD 293  
 ER+ +V+Q TINFETDLFWPIIR EK+SG Y ++ +SPKVIADHIR ++FAIGD  
 Sbjct: 232 ERKVSVIQVPIINFETDLFWPIIRATKISGTETYSGSHRADVSFRIADHIRTVPFAIGD 291  
 Query: 294 GALP+NGRGYVLRLLRLRAVHMRRLGINETFLVKLVPTVVGIMESTYPEVLEKEDFIR 353  
 GALP NBRGKYVLRLLRLRAV ++GIG+ K+Y+LVP VG IM +YPEV EK FI+  
 Sbjct: 292 GALPSNBRGKYVLRLLRLRAVRYAKQIGIDRPFMYELVFPVVGIMVDFPEVKEKAAPIQ 351  
 15 Query: 354 KIVKREESTFARTIDAGSGHLDLSLLPQLKAGKDTLEGKDIFKLYDTYGPPVLTETLAE 413  
 K+VK EEE F T++ G L+ ++ + K+EG T+ G D+FLYDTYGPPVLTET E  
 Sbjct: 352 KVVKTEERPHETINBGLSILEKVIDKAKSGASTISGSDVFLYDTYGPPVLTETEEVE 411  
 20 Query: 414 DAGYKIDBGFKSAMKEQDPAARA+VVKGGSGMGMQNETLAGIVSESRF-EYDTYSLESLI 472  
 +G ++D +GF++ M+ Q+RAR A + GSM +Q+E L I +S F Y S E++  
 Sbjct: 412 EQGLQVDLDGFSASMERQREARTARCOAGSMQVQDEVLGQITVDSTFGYKQLSTETTI 471  
 25 Query: 473 SVIADNERTETAVSEQ-ALLVFAQTPFYAEMGGQVADHGVIQNDKSDTVAEVDVQKAP 531  
 I+ D + V VQ Q A ++ +TPPYAE GGQVAD G+I+ G V V DVQKAP  
 Sbjct: 472 ETIVLDKTVADVAGQEAQVILKETPFYAESGGQVADKGIIRGANGFAV--VSDVQKAP 529  
 Query: 532 NQQLHVTNVN-ASLSVGTNYTLEINKERRLAVERNHTATHLLHAALHNVIGEHATQAGS 590  
 NQQLHVT V +L V ++ R +KNHTATHLLH AL +V+GEH QAGS  
 Sbjct: 530 NQQLHVTIVKSGTLQVNDQVQAVSETERSGIVKNHTATHLLHRAKLDVIGEHVHQAGS 589  
 30 Query: 591 LNNEEPLRFDFTHFEAVSNELRHIEQEVNEQIGVNDLTITTTTDTVETAKEMGAMALFGE 650  
 L EE LRPFDFHF V+EE IS+ VNE+IW ++ + ++AK +GAMALFGE  
 Sbjct: 590 LVSEERLRPFDFSHFQGVDEEKEKIERIVNEKIWQAKVNISTITLDEAKIAGMALFGE 649  
 35 Query: 651 KYGKVRVVOIGNYSVELQSGTHLANSSEIGLFKIVKEBGSIGSGTRRIIAVTGRQAFAY 710  
 KYG +VRVV++G+YS+ELQGG H+ N+SEIGLFKIV E GIG-G RRI AVTG++AF  
 Sbjct: 650 KYGDIVRVVEVDYSIELQGGCHVTNTSEIGLFKIVSESGIGAGVRIEAVTGKEAFLFM 709  
 40 Query: 711 RNQEDALKEIAATVIAQPLQDAARKVQALSDSLRLOKENVELKEKAAAAAGDVFKDIQ 770  
 Q D LKE AATVKA +KD +V+AL +R+LQ+EN L K AG + ++Q  
 Sbjct: 710 AKQLDLLKETAAATVIAKIVKVPVRVEALQQQIRELQRENESLNKLGNMESGLVNEVQ 769  
 45 Query: 771 EAKGVRFIASQVDVADAGALRTFADNWKQKDYSDVLVLVAALIGKMNVLVASKTKDV--- 827  
 +GV +A + AD LR+ D KQ+ S +VL A KVN+ VA TID+  
 Sbjct: 770 KIEGVPLAKAISGADMGLRSIVDKLQETPSVIVLGTASBGKNI-VAGVTKDLINK 828  
 Query: 828 --HAGNIRGLAPIVAGRGQKPDMAAGQSDASKIALLAAVE 870  
 HPAK ++K +A G GGG+PDMA AGG K+ + L+ V E  
 50 Sbjct: 829 GYHAGKLVKEVATRCGGGGGRPDMAAGGKQPEKLDQALSFPYIE 873

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4505> which encodes the amino acid sequence <SEQ ID 4506>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.41 Transmembrane 805 - 821 ( 804 - 822)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2763 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 862/870 (99%), Positives = 864/870 (99%)

Query: 1	MKELSSAQIRQWLDQFWKSNHGVSPSANLVFVNDPTLLKINSQVATLKCCYDQGSVIPEN	60	
Sbjct: 1	MKELSSAQIRQWLDQFWKSNHGVSPSANLVFVNDPTLLKINSQVATLKCCYDQGSVIPEN	60	
5	Query: 61	PRITNAQKSIRINDIENVGKTAHHMFEMLGNFSIGDYFRDEAIENWGFELLTSPWPDF	120
Sbjct: 61	PRITNAQKSIRINDIENVGKTAHHMFEMLGNFSIGDYFRDEAIENWGFELLTSPWPDF	120	
10	Query: 121	PKDKLWMTYYPDDKDSYNRWIACGVPSHLVPIEDNFWEIGAGPSGPDTRIFFDRGKDFD	180
Sbjct: 121	PKDKLWMTYYPDDKDSYNRWIACGVPSHLVPIEDNFWEIGAGPSGPDTRIFFDRGKDFD	180	
15	Query: 181	PENIGRLRLAEDIENDRYTEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAY	240
Sbjct: 181	PENIGRLRLAEDIENDRYTEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAY	240	
20	Query: 241	MQGAKTNFETDLFMPILIREVEKLSGKTYDPDGNMSPKVIADHIALGFAIGDALPQNE	300
Sbjct: 241	MQGAKTNFETDLFMPILIREVEKLSGKTYDPDGNMSPKVIADHIALGFAIGDALPQNE	300	
25	Query: 301	GRGYVLRLLRRVNHGRRLGINETFLYKLVPTVQIMBESYTPVLEKRPDIKIVKREE	360
Sbjct: 301	GRGYVLRLLRRVNHGRRLGINETFLYKLVPTVQIMBESYTPVLEKRPDIKIVKREE	360	
30	Query: 361	ETPARTIDAGSGHLSLQAQKABGKOTLEKGDIFKLYDTPGFVPELTELBAEDAGYKID	420
Sbjct: 361	ETPARTIDAGSGHLSLQAQKABGKOTLEKGDIFKLYDTPGFVPELTELBAEDAGYKID	420	
35	Query: 421	HGFKSAMKEQQDARAANVKGSGMQMNETLAGIVEBSRFEDYTSLESSLSVIADNE	480
Sbjct: 421	HGFKSAMKEQQDARAANVKGSGMQMNETLAGIVEBSRFEDYTSLESSLSVIADNE	480	
40	Query: 481	RTEAVSBQALLVFAQTFFPABMGQVADGHVIGKDKGTVAEVDVQKAPKQPLHTVN	540
Sbjct: 481	RTEAVSBQALLVFAQTFFPABMGQVADGHVIGKDKGTVAEVDVQKAPKQPLHTVN	540	
45	Query: 541	VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNVIQSHATQAGSLNEESPLRFD	600
Sbjct: 541	VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNVIQSHATQAGSLNEESPLRFD	600	
50	Query: 601	PTHFEAVSNEELRHIEQVNEQIHNDLTITTTTDOVETAKEMGAWALPGEKYGVVRVQ	660
Sbjct: 601	PTHFEAVSNEELRHIEQVNEQIHNDLTITTTTDOVETAKEMGAWALPGEKYGVVRVQ	660	
55	Query: 661	IGNYSVELCGTHLANSSEIGLFIKIVKEGIGSGTERRIAVTRQAFAYRNQEDALKEI	720
Sbjct: 661	IGNYSVELCGTHLANSSEIGLFIKIVKEGIGSGTERRIAVTRQAFAYRNQEDALKEI	720	
60	Query: 721	AATVKAPQLKDAARKVQALSLSRLDLQKNVLEKKAIAAAGGVFDQKQGVVFAS	780
Sbjct: 721	AATVKAPQLKDAARKVQALSLSRLDLQKNVLEKKAIAAAGGVFDQKQGVVFAS	780	
65	Query: 781	QVDVADAGALRTFADNWKQDYSDVLVLAAGIKGVNVLVASKTIDVHAGNMIKIAPIV	840
Sbjct: 781	QVDVADAGALRTFADNWKQDYSDVLVLAAGIKGVNVLVASKTIDVHAGNMIKIAPIV	840	
70	Query: 841	AGROGGKPMAMAGSDASKIAELAAVAE 870	
Sbjct: 841	AGROGGKPMAMAGSDASKIAELAAVAE 870		

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1620-

**Example 1467**

A DNA sequence (GBSx1553) was identified in *S.agalactiae* <SEQ ID 4507> which encodes the amino acid sequence <SEQ ID 4508>. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2974 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9747> which encodes amino acid sequence <SEQ ID 9748> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP: CAB15920 GB: Z99123 yxjI [Bacillus subtilis]
Identities = 42/144 (29%), Positives = 73/144 (50%), Gaps = 2/144 (1%)

Query: 17 IKERNPSLGGKFTITDLTGLPCYHVEGSLFPLPKTFKVFDEEHLISQIEKKVLSFLPKF 76
      +K+KNFS F I D + VEG F L + ++ D + IE+K++S LP++
Sbjct: 6 MQKNPSFKDAFHIYDRDEQETFKVEGRPFSLGDSLQMTDSSGKTLVSIQKLSLLPRY 65

Query: 77 NVTLANGNHFTIKKDFSLKPHYTIKDLMEVKGNGFWMDPQLLKDNQVIANISQQWFRM 136
      +++ + K +F KP + I L+ E+ G+ W +PQL V ++S++W
Sbjct: 66 EISIOGKTVCVEVTKVTFSPKPKFVISGLNWEIDGDLWRDEFQLTDGENVMVSVKLOLSW 125

Query: 137 TSTYQVEVYSETYNDLTISLWIAI 160
      +Y +++ E D+ I IAI
Sbjct: 126 GDSYHLOIAYE--EDVLICTAI 147

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1468**

A DNA sequence (GBSx1554) was identified in *S.agalactiae* <SEQ ID 4509> which encodes the amino acid sequence <SEQ ID 4510>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3833 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP: BAA36674 GB: AB016282 ORF17 [bacteriophage phi-105]
Identities = 45/133 (33%), Positives = 74/133 (56%), Gaps = 5/133 (3%)

Query: 2 RYTYLALFVFDKINGGYNISFPDHFCAFSZADGIAKAI PNAREVLEIYTIMFDEGKEFP 61
      +Y Y ALF+ D + G ++FED G + +S RA+ A+E + ++ FR +G P
Sbjct: 5 RYIYPALFDYDDO--GITVTFPDLPGCITFGNSGCHALTMAKENAMALHLYGFPDQCDIIP 62

Query: 62 KASSPKALASNLASDEDVIVQISVDTLVRHRSERSKI VNKTVTLPSWLVEVGKENVKVS 121
      +A+ K + A + + I R + V KI+T+P W+ ++ KE+KVN+G
Sbjct: 63 EATPSKEIK---AESQGVVLLETWMPFPRHIMENAAVXKLTITIPRWMDIAKHVNVYS 119

```

-1621-

Query: 122 QLLQKAIREEQLV 134  
 QLLQ+AI+E L +  
 Sbjct: 120 QLLQEATKEHLGI 132

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1469

- 10 A DNA sequence (GBSx1555) was identified in *S.agalactiae* <SEQ ID 4511> which encodes the amino acid sequence <SEQ ID 4512>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1484 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:BAA25696 GB:AB010712 NADH oxidase/alkyl hydroperoxidase  
 reductase [Streptococcus mutans]  
 Identities = 383/509 (75%), Positives = 441/509 (86%)
- 25 Query: 1 MVLDEIKIAQLAQLYLDLLESDIVLQADLGDNDNSQKVKFLDEIVAMSDRISLSTHLKR 60  
 M ID BIK QL QYL LLES+IVLQA L D+ NSQKVK+FL EIVAMS ISLE L R  
 Sbjct: 1 MALDAIEIKELQQLYQLLLESETIVLQQLKDANSQKVEFLQIEIVAMSPMISLSEKELPR 60
- 30 Query: 61 QPFGIAKKGHESRVIFSGPLMGHEFTSFILALLQVSGRAPKVDIEDIKRIRGIETINL 120  
 PSF IAKNG ES V F+GLP+GHEFTSFILALLQVSGR PKV+ DI+KRI+ +++ ++  
 Sbjct: 61 TPGFRIAKKGQESGVFEAGLPLGHEFTSFILALLQVSGRPPKVTEDIVKRIQAVDEPMHF 120
- 35 Query: 121 ETVVSLTCHNCDFVVQAFNIMAVLNFNITHIMIEGGMQDEVKSGKIMSVEFTYKDEEF 180  
 ETVVSLTCHNCDFVVQAFNIM+V+NENI+HTM+EGGM++DE+++KGIMEVFTYKDE EF  
 Sbjct: 121 ETVVSLTCHNCDFVVQAFNIMSVVFNISHTMVEGGMFKDEIEAKGIMSVEFTYKDGTEF 180
- 40 Query: 181 TSGRATIEQLLQDGLDGLDAEAFADGKGVYDVLVIGGGPAGNSAITYAARKGLKCTGILAE 240  
 TSGRA+IEQLL+ + GPL +AF DKG+DVLVIGGGPAGNSAITYAARKG+KTG+LAET  
 Sbjct: 181 TSGRASIEQLLQDGLDGLDGLDAEAFADGKGVYDVLVIGGGPAGNSAITYAARKGVKTGLLAET 240
- 45 Query: 241 FGGQVETVGIENMIGTLYTGGKLMQAIIEHTKSYDIDIKSQLANGIEKKELVEVTLA 300  
 GGQV+ETVGIENMIGT Y EGP+IMAQ+EEHTKSY +DI+K+ A I+K +LVEV L  
 Sbjct: 241 MGGQVMETVGIENMIGTPYVGGPQMAQVEEHTKSYDVIDMKAPRAKSIQKTDLVEVLD 300
- 50 Query: 301 NGAILQAKTAILALGAKWRINIVGSEEFNKGVTYCPHCXGLPEFGKDVAVIGGGNSGM 360  
 NGA L+AKTA+LALGAKWR INVEGE+EF NKGVTYCPHCXGLF K VAVIGGGNSG+  
 Sbjct: 301 NGAILKAKTAVIALGAKWRKINIVGSEEFNKGVTYCPHCXGLFETKRVAVIGGGNSGL 360
- 55 Query: 361 EAALDLAGVTKHVTVLEFLPELKADQVLQERAAKTDNLTIKKNVATKIDVGHEDVTGLNY 420  
 EAA+DLAG+ HV +LEFLPELKAD++LQ+RA DN+TIL NVATK+I+G DRY GL V  
 Sbjct: 361 EAIDLGLAGSHVYILEFLPELKADKILQRAKALDNTILTNVATKIEIGNDVRHGRY 420
- Query: 421 TDRDTNEEKHIDLEGVVFQIGLVSTSWLKDSGIELNREQEIVDKPGSTNIPGIFAA 480  
 +DR TNEE +DLGCVVFQIGLVSTSWLKDSG+LNR+ EI+V K G+TNIP IFAAGD  
 Sbjct: 421 SDRTTNEEYLLDLEGVVFQIGLVSTSWLKDSGLALNKEGELIVAKDGATNIPALFAAGD 480
- Query: 481 CTDAAYKQIIISMGSGATAALGAFDYLLIR 509  
 CTD+AYKQIIISMGSGATAA+GAFDYLLIR  
 Sbjct: 481 CTDASAYKQIIISMGSGATAALGAFDYLLIR 509

-1622-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4513> which encodes the amino acid sequence <SEQ ID 4514>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0654 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 419/510 (82%), Positives = 472/510 (92%)

Query: 1 MVLDEKIKQAQLQYLLESDIVLQADLADNDNSQKVDFLDEIVAMSDRISLESTHLKR 60
      1 M L +IK QLAQYL LLE+D+VLQ LGDN+ SQKVKDF++EI AMS+RIS+R+ L R
Sbjct: 1 KALSFDIKKQAQLQYLLEADLVLQVSLGDNESQKVKDFVDEIAMSERISLENTILDR 60

Query: 61 QPSFGTAKKGHSRVSFQSLPMGHEFTSFILALLQVSGRAFKVDDEIKRIKGIETINL 120
      61 QPSF +AKKGH S V+P+GLP+GHE TSFILALLQVSGRAFKVD+D+I RIK I++ ++
Sbjct: 61 QPSFKVAKKGHSGSVFAGLPLGHELTSTFILALLQVSGRAFKVDQVDIDRIKALDRPLH 120

Query: 121 ETVVSLTCHNCPTDVQAFNIMAVLNPNITHIMWEGMYQDEVKSGIMSVPTVYKQDEEF 180
      121 ETVVSLTCHNCPTDVQA NIM+VLN I+HIM+EGM+QDEVK+SGIMSVPTV+ D EEF
Sbjct: 121 ETVVSLTCHNCPTDVQALNIMSVLNDKISHTVVEQGMFQDEVKSGIMSVPTVFLDQEEF 180

Query: 181 TSGRATTIEQLLEQLDGLDAEAFADKGVYDVLVIGGGPAGNSAAIYAARKGLKTGLAET 240
      181 TSGRATTIEQLLEQ+ GPL EAFADK+YDVLVIGGGPAGNSAAIYAARKGLKTGLAET
Sbjct: 181 TSGRATTIEQLLEQLAGPLSEAFADKGLYDVLVIGGGPAGNSAAIYAARKGLKTGLAET 240

Query: 241 FGGQVETVGIENMIGTLYTEGPKLMAQIEEHTKSYDDIHKQLATGIEKKELAEVTLA 300
      241 FGGQV+ETVGIENMIGTLYTEGPKLMA++E HTKSYD+DIK+QLAT IEKKE +EVTLA
Sbjct: 241 FGGQVMEVGIENMIGTLYTEGPKLMAEVAHTKSYDVIDIKAQLATSEKKEHIEVTLA 300

Query: 301 NGAILQAATAILALGAKWRNINVPGEIEPRNKGVITYCPHCDGLPEKGDVAVIQQGNSGL 360
      301 NGA+LQAATAILALGAKWRNINVPGE+EPNKGVTYCPHCDGLPEKGDVAVIQQGNSG+
Sbjct: 301 NGAVLQAATAILALGAKWRNINVPGEDEPRNKGVITYCPHCDGLPEKGDVAVIQQGNSGL 360

Query: 361 EAALDLAGVTKHVTVLEFLPELKADQVLQRAAKTDNLTILKNVATKDIVGEDRVITGLNY 420
      361 EAALDLAG+ KHV VLEFLPELKAD+VLQ+RAAKT+N+TI+KNVATKDIVGEDRVITGLNY
Sbjct: 361 EAALDLAGLKHVVVLEFLPELKADQVLQRAAKTNMTTIKNVATKDIVGEDRVITGLNY 420

Query: 421 TDRDNEEKHIDLEGVVFQVIGLVPTSNLWDSGIELMRQEVVDKFGSTNIPGIPAAGD 480
      421 T+R+D +E+K+DLRGVVFQVIGLVPT+T+NLWDSG+ L +R EI+VDK GSTNIPGIPAAGD
Sbjct: 421 TERDSECKHLDLEGVVFQVIGLVPTANLWDSGVNLTRGEIIVDKHGSTNIPGIPAAGD 480

Query: 481 CTDAAYKQIIISMGSATAAICAFDYLIQ 510
      481 CTD+AYKQIIISMGSATAAICAFDYLIQ
Sbjct: 481 CTDSAYKQIIISMGSATAAICAFDYLIQ 510

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1470

A DNA sequence (GBSx1556) was identified in *S.agalactiae* <SEQ ID 4515> which encodes the amino acid sequence <SEQ ID 4516>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2906 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```



-1623-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA25695 GB:AB010712 alkyl hydroperoxidase (Streptococcus mutans)  
Identities = 167/186 (89%), Positives = 179/186 (95%)

Query: 1 MSLVGKEIIEFSAQAYHDGKFITVTNEDVKGKMAVFCFPADPSFVCPTELGLDQEQYET 60  
MSLVGKE++EFGAQAYH G+F+TV NEDVKGKMAVFCFPADPSFVCPTELGLDQEQY T  
Sbjct: 1 MSLVGKEMVFEFSAQAYHGKFFVTNNESDVKGKMAVFCFPADPSFVCPTELGLDQEQYAT 60

Query: 61 LKSLDVEVYSVSTDTTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQGFVDLQDGLAQRG 120  
L+SL VEVYSVSTDTTHFVHKAWHDDSDVVGTTITY MIGDPSH++SQGF+VLG+DGLAQRG  
Sbjct: 61 LQSLDVEVYSVSTDTTHFVHKAWHDDSDVVGTTITYPMIGDPSHLSQGFVEVLGSDGLAQRG 120

Query: 121 TFIIDPDGVIQMMBINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKBAETLTPSL 180  
TFI+DPDG+IQMME+NADGIGRDASTLIDKVRAAQYIRQH GEVCPAKWKBAETL PSL  
Sbjct: 121 TFIIDPDGIIQMMBINADGIGRDASTLIDKVRAAQYIRQHPGEVCPAKWKBAETLTPSL 180

Query: 181 DLVGKI 186  
DLVGKI  
Sbjct: 181 DLVGKI 186

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4517> which encodes the amino acid sequence <SEQ ID 4518>. Analysis of this protein sequence reveals the following:

Possible site: 15  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3022 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/186 (93%), Positives = 181/186 (97%)

Query: 1 MSLVGKEIIEFSAQAYHDGKFITVTNEDVKGKMAVFCFPADPSFVCPTELGLDQEQYET 60  
MSL+GKEI EFGAQAYHDGKFITVTNEDVKGKMAVFCFPADPSFVCPTELGLDQEQYET  
Sbjct: 1 MSLVGKEIAEFSAQAYHDGKFITVTNEDVKGKMAVFCFPADPSFVCPTELGLDQEQYET 60

Query: 61 LKSLDVEVYSVSTDTTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQGFVDLQDGLAQRG 120  
LKSL VEVYSVSTDTTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQ F+VLG+DGLAQRG  
Sbjct: 61 LKSLDVEVYSVSTDTTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQGFVEVLGSDGLAQRG 120

Query: 121 TFIIDPDGVIQMMBINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKBAETLTPSL 180  
TFI+DPDG+IQMMEBINADGIGRDASTLIDK+ AAQY+R+H GEVCPAKWKBAETLTPSL  
Sbjct: 121 TFIIDPDGIIQMMBINADGIGRDASTLIDKTHAAQYVRKHGEVCPAKWKBAETLTPSL 180

Query: 181 DLVGKI 186  
DLVGKI  
Sbjct: 181 DLVGKI 186

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1471**

A DNA sequence (GBSx1557) was identified in *S.agalactiae* <SEQ ID 4519> which encodes the amino acid sequence <SEQ ID 4520>. This protein is predicted to be 30S ribosomal protein S2 (rpsB). Analysis of this protein sequence reveals the following:

Possible site: 60

-1624-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.4462 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA50276 GB:X70925 30S ribosomal protein [Pediococcus  
           acidilactici]  
       Identities = 190/260 (73%), Positives = 226/260 (86%), Gaps = 4/260 (1%)  
       Query: 1 MAVISMKQLLEAGVHPGHQTRRWNPIMAKYIFTERNGIHEVIDLQQTVKLADQAYEFPVDA 60  
               M+VISMKQLLEAGVHPGHQTRRWNPIM +IFTERNGI++IDLQ+TVKL D AY FV+D  
       15 Sbjct: 1 MSVISMKQLLEAGVHPGHQTRRWNPIMKPFIFTERNGIYIIDLQQTVKLIDNAYNPFVDV 60  
       Query: 61 AANDAVILFVGTKKQAAEVAEAKRAGQYFINHRWLGOTLTNNGTIQKRIARLKEIKRM 120  
               AAND V+L FVGTKKQA A+ ESAKRAGQ++NHRWLGOTLTNN TIQRI RLK++K+M  
       20 Sbjct: 61 AANDGVTLFVGTKKQAQTAEBAKRAGQYFNHRWLGOTLTNNNTIQKRIKRLKDLGKN 120  
       Query: 121 EEBGTFELLPKKEVALLNKQARLEKFLGIEDMPRIPOVMYVVDPHKEQIAVKEAKKLG 180  
               EE+GTF+ LPKKEVALLNKQ+ +LEKFLGIEDMP IDPV++VYDP KEQIA+KEA+KL  
       25 Sbjct: 121 EEDGTFDRLPKKEVALLNKQKDLKFLGIEDMPHIDPVLVVDPRKEQIAIEAKQKLN 180  
       Query: 181 IPVVMVDTNADPDDIDVILPANDAIRAVKLITSLADAVIEGRQGEDADV---DPAQ 236  
               IPVVMVDTIN DDP +DVIIIP+NDADIRAV+LITSK+ADAV+ESRQGED + +A+  
       30 Sbjct: 181 IPVVMVDTINTDPQVDVIIIPGNDADIRAVRLITSKNADAVIEGRQGEDAEVQGEVAE 240  
       Query: 237 EAQADSIEIEVVEGSGND 256  
               DS+E++ + VE +N+  
       30 Sbjct: 241 GVSXDSLEDLKITVERGSGNE 260

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4521> which encodes the amino acid  
 sequence <SEQ ID 4522>. Analysis of this protein sequence reveals the following:

35       Possible site: 60  
       >>> Seems to have no N-terminal signal sequence  
       ----- Final Results -----  
       40           bacterial cytoplasm --- Certainty=0.4462 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

          Identities = 241/254 (94%), Positives = 248/254 (96%)  
       45       Query: 1 MAVISMKQLLEAGVHPGHQTRRWNPIMAKYIFTERNGIHEVIDLQQTVKLADQAYEFPVDA 60  
               NAVISMKQLLEAGVHPGHQTRRWNPIMAKYIFTERNGIHEVIDLQQTVKLADQAYEFPVDA  
               Sbjct: 1 MAVISMKQLLEAGVHPGHQTRRWNPIMAKYIFTERNGIHEVIDLQQTVKLADQAYEFPVDA 60  
       50       Query: 61 AANDAVILFVGTKKQAAEVAEAKRAGQYFINHRWLGOTLTNNGTIQKRIARLKEIKRM 120  
               AANDAVILFVGTKKQAAEVA+EA RAQGYFINHRWLGOTLTNNGTIQKRIARLKEIKRM  
               Sbjct: 61 AANDAVILFVGTKKQAAEVAEADTRAGQYFINHRWLGOTLTNNGTIQKRIARLKEIKRM 120  
       55       Query: 121 EEBGTFELLPKKEVALLNKQARLEKFLGIEDMPRIPOVMYVVDPHKEQIAVKEAKKLG 180  
               EEBGTF+L+LPKKEVALLNKQARLEKFLGIEDMPRIPOVMYVVDPHKEQIAVKEAKKLG  
               Sbjct: 121 EEBGTFDVLPKKEVALLNKQARLEKFLGIEDMPRIPOVMYVVDPHKEQIAVKEAKKLG 180  
       60       Query: 181 IPVVMVDTNADPDDIDVILPANDAIRAVKLITSLADAVIEGRQGEDADVDFAEQA 240  
               IPVVMVDTNADPDDID+IIPANDAIRAVKLIT+KLADA+IEGRQGEDADVDF+QA  
       60 Sbjct: 181 IPVVMVDTNADPDDIDIIIPANDAIRAVKLITAKLADAIEGRQGEDADVDFADTQA 240  
       Query: 241 DSIEIEVVEGSGN 254  
               DSIEIEVVEG N  
       Sbjct: 241 DSIEIEVVEGSDN 254

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1472

- 5 A DNA sequence (GBSx1558) was identified in *S. agalactiae* <SEQ ID 4523> which encodes the amino acid sequence <SEQ ID 4524>. Analysis of this protein sequence reveals the following:

Possible site: 49  
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2648 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE73435 GB:AL139077 elongation factor TS [Campylobacter jejuni]  
Identities = 169/358 (47%), Positives = 226/358 (62%), Gaps = 19/358 (5%)

- 20 Query: 1 MAEITAKLVKELREKSGAGVMDAKKALVETDGDLDKATLREKGMKAAKKADRVAAEG 60  
M EITA +VKEELRE +GAG+MD K AL ET+GD DKA++LLREK+ KAACKADR+AAEG  
Sbjct: 1 MTEITAAVMKELREKSGAGVMDCKNALSETNGDFKAVQLLREKGLGKAACKADRLAAEG 60

- Query: 61 LTGVVY--DGNVAAVIEWNAETDFVAKNDQFVTLVNETAKVIABGRPSINBEALALINFS 118  
L V V D A V E+N+ETDFVAKNDQF+ L +T I + EE +T+ +  
25 Sbjct: 61 LVSVKVSDDFTSATVSEINSETDFVAKNDQFIALTKDTTAHIQSNLSQSVBELHSSTI-N 119

- Query: 119 GETLSCAFVTATATIGEKISFRFALVEKTOBOHFGAYQHNGRIGVITV-----VEG 171  
G E+ + ATIGE + RRFA ++ Y H GR+GV+ V  
30 Sbjct: 120 GVKFEYELKQIATIGENLVRRFATLKAGANGVVGVIYHTNGRVGVVIAACDSAEVAS 179

- Query: 172 GDDALAFQVSMHVAAKPTVLSYTELDAQFVHDELAQLAHKTEBQDNESRAV---NKEAL 228  
L +Q+ MH+AAH+P+ LSY +LD FV +E L +E+ANE R + NKP  
35 Sbjct: 180 KSRDLRLQICMLIAAMRPSYLSYEDLMTTFVENEYKALVBLEKENEERRRLKDPKPFH 239

- Query: 229 PFLKYGSKAQLTDEVLAQAEDIKAEALAEKPKPKINDKIIVFGKMDRFLDNTKVQCEYT 288  
++ S+ QL+D ++ +AEE IK EL A+GKPKIWD I+PKM+ F+ DN++D + T  
40 Sbjct: 240 KIPQFARQKQLSDAILKGAEEKKEELKQKPKIKINDIIPKMGNSFIADNSQLDSILT 299

- Query: 289 LLAQVYIMDDGKTVZAYLESV-----NAKAVAFVRFVEVGBGLEKASNDFAEVAATH 340  
L+ Q Y+ND KIVE + K V P+ FEVGB+EK + DF ABVA +  
45 Sbjct: 300 LMSQFVYMDDKTVEVQVIAEKEKEFGGKIKIVEFICFEVGBGLEKKTDFAAEVAQL 357

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4525> which encodes the amino acid sequence <SEQ ID 4526>. Analysis of this protein sequence reveals the following:

- 45 Possible site: 38  
>>> Seems to have no N-terminal signal sequence  
----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3942 (Affirmative) < succ>  
50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 307/344 (89%), Positives = 327/344 (94%)  
55 Query: 1 MAEITAKLVKELREKSGAGVMDAKKALVETDGDLDKATLREKGMKAAKKADRVAAEG 60  
MAEITAKLVKELREKSGAGVMDAKKALVETDGD+DKA+ELREKGMKAAKKADRVAAEG  
Sbjct: 33 MAEITAKLVKELREKSGAGVMDAKKALVETDGDMDKAVELREKGMKAAKKADRVAAEG 92

-1626-

Query: 61 LTGVYVGNVAARVAVVNAETDFVAKNDQFVTLVNSTAKVIAEGRPSNNREALALTPPSGE 120  
 LTGVYV GNVAARVAVVNAETDFVAKN QV LVN TARVIAERG+P+NN+BAIALI MPSPGE  
 Sbjct: 93 LTGVYVGNVAARVAVVNAETDFVAKNAQFVELVNAATKRVIAEGRPANDEALALTPPSGE 152

5 Query: 121 TLSCAFVTTATATIGEKISFRFPALVEKTBQHFQAYQHNGRIGVITVVBGGDALARQV 180  
 TL +A+V ATATIGEKISFRFPAL+EK DEQHFQAYQHNGRIGVI+VVBGGDALARQV  
 Sbjct: 153 TLAEAYVNAATATIGEKISFRFPALTEKADQHFQAYQHNGRIGVISVVBGGDALARQV 212

10 Query: 181 SMHVAAMKPTVLSYTELDAQFVHDELAQINHKIKQDNEBRAMVKNPAPFLKYGSKAQLT 240  
 SMH+AAAMKPTVLSYTELDAQF+ DELAQINH IR DNEBRAMV+KPAFLPKYGSKAQL+  
 Sbjct: 213 SMHIAAMKPTVLSYTELDAQFIKDELAQINHAILEDNEBRAMVDKPAFLPKYGSKAQLS 272

15 Query: 241 DEVIAQAEEDIKAEAAAGKPEKIMDKIVPGKMRPMLNTKVDQGYTLAQVYIMDDSK 300  
 D-VI AR DIKAEAAAGKPEKIMDKI+PGKMRPMLNTKVDQ YTLAQVYIMDDSK  
 Sbjct: 273 DDVITAAEADIKAEAAAGKPEKIMDKIIPGKMRPMLNTKVDQAYTLAQVYIMDDSK 332

20 Query: 301 TVEAYLESVNAKAVAFVRFEVGGIEKASNDFAEVAATMAAAL 344  
 TVEAYL+SVNAKA+AF RFEVGGIEK +NDFE+KVAATMAAAL  
 Sbjct: 333 TVEAYLSVNAKAIAFARFVEVGGIEKKAANDFSEVAATMAAAL 376

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1473

A DNA sequence (GBSx1559) was identified in *S.agalactiae* <SEQ ID 4527> which encodes the amino acid sequence <SEQ ID 4528>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1312 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1474

A DNA sequence (GBSx1560) was identified in *S.agalactiae* <SEQ ID 4529> which encodes the amino acid sequence <SEQ ID 4530>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood = -7.86 Transmembrane 128 - 144 ( 124 - 152)  
 INTEGRAL Likelihood = -4.57 Transmembrane 35 - 51 ( 33 - 53)  
 INTEGRAL Likelihood = -4.04 Transmembrane 92 - 108 ( 87 - 111)

50 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4142 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAH04953 GB:AP001511 small multidrug export related protein  
 [Bacillus halodurans]



-1628-

**Example 1475**

A DNA sequence (GBSx1561) was identified in *S. agalactiae* <SEQ ID 4531> which encodes the amino acid sequence <SEQ ID 4532>. This protein is predicted to be CtsR protein (ctsR). Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.3672 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CA891548 GB:AJ249133 CtsR protein [Lactococcus lactis]
   Identities = 74/146 (50%), Positives = 103/146 (69%), Gaps = 3/146 (2%)

   Query: 4 KNTSDNIREYIKSLLEQSGIAEIKRNLADTFQVVFSGQINVIKTRFTESRGYVVSERG 63
           KNTSD IE Y++ LLE++ + EIKR++L+ F VVFSGQINVIKTRFT S+G+ VESKRG
   Sbjct: 5 KNTSDIIEYIKSLLEQSGIAEIKRNLADTFQVVFSGQINVIKTRFTASKGDFVESKRG 64

20  Query: 64 GGGYIRIAKVHPSDQHQLFGNMLSTIGERISQVFDLLQLLFDSEIITEREGNLLIATS 123
           GGGYI+I K +S +H+ + + +S + D++QLLFDE+++TEREGNL+L
   Sbjct: 65 GGGYIKIKVQYSARHEFLTALYQKVFANLSKFAHDIVQLLFDEKVLITEREGNLLIIVI 124

25  Query: 124 GDOVLGEQASVIRARMLRKLQLRDLR 149
           D G + R M++ ++ RLDR
   Sbjct: 125 TD---GAISPFPTGIMMSIINRLDR 147

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4533> which encodes the amino acid sequence <SEQ ID 4534>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
35  bacterial cytoplasm --- Certainty=0.2514 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40  Identities = 117/151 (77%), Positives = 131/151 (86%)

   Query: 1 MAIKNTSDNIREYIKSLLEQSGIAEIKRNLADTFQVVFSGQINVIKTRFTESRGYVES 60
           M KNTSD+IREYIK LL +SGIAEIKRS LAD+QVVFSGQINVIKTRFTESRGY VES
   Sbjct: 1 MPTKNTSDSIREYIKLLAKSGIAEIKRNLADTFQVVFSGQINVIKTRFTESRGYVES 60

45  Query: 61 KRGGGGYIRIAKVHPSDQHQLFGNMLSTIGERISQVFDLLQLLFDSEIITEREGNLLI 120
           KRGGGGYIRIAKVHPSD+H L GN++TY + ISSQVF D IQLLFDE ++TEREGN+IL
   Sbjct: 61 KRGGGGYIRIAKVHPSDKIHILGIMLMTIEDCISRSQVPTDSIQLLFDSEIITEREGNII 120

50  Query: 121 ATSGDDVLGEQASVIRARMLRKLQLRDLRKG 151
           A + DDVLG S IRARML+LLQR+DRKG
   Sbjct: 121 AVASDDVLGTDGSTRARMLRKLQLRDLRKG 151

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1476

A DNA sequence (GBSx1562) was identified in *S. agalactiae* <SEQ ID 4535> which encodes the amino acid sequence <SEQ ID 4536>. This protein is predicted to be ClpC (clpB-1). Analysis of this protein sequence reveals the following:

5 Possible site: 49  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.34 Transmembrane 32 - 48 ( 32 - 49)  
 10 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1935 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AD01783 GB:AF023422 ClpC (Lactococcus lactis)  
 Identities = 401/831 (48%), Positives = 571/831 (68%), Gaps = 52/831 (6%)  
 Query: 4 YSIKLQEVFR:LAQFOAARYSHYLESWHLLAMVLVHDSVAGLTFAYE---SEVAIEY 60  
 Y+ L +F A A +Y+ +ES HLL AM S+A A S++ I+  
 20 Sbjct: 8 YTPILDRIFEKAAEYAHQYQYGTIESAEHLAAMATTSGSIAYSILAGMNVDSDDLIDLE 67  
 Query: 61 EAATILALGRAPKEITNYQFLEQSPALKIKILKARNISIVVGAEDVGRHVLAMLVNK 120  
 ++ ++ R+ L SP ++++ +A +++ AE VGTEH+L A+L ++  
 25 Sbjct: 68 DLSSHVKVKRSE-----LRFSPRAEEVTVASFLAVHNNAAVGTREHLLYALQLQE 118  
 Query: 121 DLLATRIELEIVGFRGQDDGESVRMVLRLKALERHAGF-TKDDIKAIYELRNPKKAKSGAS 179  
 D ++L+L ++ +V LRK +E+ G ++ KA+ + K K A  
 Sbjct: 119 DGFLQLLLK-----QKINIVSLRKEIKRTGLVLPENKAVTPMSKRKNKAKGVAS 169  
 30 Query: 180 FSDMKKPPSTAGDLADFTDLDSQMAVDGEIEFVIGRDKIESRMVQLSRKTNKPNVLVGD 239  
 S+ L + DL++ A G++P+IGR+ E+ R++ +LSR+TKQNPVLVGA  
 Sbjct: 170 -----NSSPTLDSVSSDLTEAARSGKLDPMIGREAEVDRLIHLSRTRKNPNVLVGE 222  
 35 Query: 240 AGVGKTALAYGLAQRIANGNIPIYELRDMVLELDMMSVAGTRFRGDFEERNQIADIE 299  
 GVKG+A+ GLAQRI NG +P L + R++ L+M +VVGAT+FRG+FE+R+ I+ ++  
 Sbjct: 223 PGVGKSAITIEGLAQRIVMGQVPIGLANSRINALNMATVVGATKFRGEFEDRLTAIVEVS 282  
 Query: 300 EDGHIIIFIDELHTIMGSGSGIDSTLDANMLKPALARGTLRTVGATTQERYQKHIEKDA 359  
 D +I+FIDELHTI+G+G G+DS DANMLKPALANG +VGATT EYQK+IERD  
 40 Sbjct: 283 ADPDVIIIFIDELHTI+GAGGMDGVNDAANMLKPALARGDPQMVGATTYHEVQKYTEKDE 342  
 Query: 360 ALSRRFAKVLVEEPNEIDAYEILLGLKPAYEAFHNTVDSAEVMTAKVVAHRYLTISNLP 419  
 AL RR A++ V+EP+ +A+ IL GL+ +E +H V +D+A+ +AV ++ RY+TS+ LP  
 Sbjct: 343 ALERRLARINVDERSFDEAIALQLGRKPKEDYHQVQKPTDQAIKSAVTLVRYMTRSKLP 402  
 45 Query: 420 DSAIDLLDEASATVCMNMKKNAPSLLT-----EVDQAILDDDMKSA----- 460  
 D AIDLLDEA+ V++++K ++ E+ +A++ D+K++  
 Sbjct: 403 DKAIDLLDEAARVILLKTKQKQNVFELEDPVKAQKELRAVILKLDVKSRIKEKAVEK 462  
 50 Query: 461 --SKALKASYGKKRKPPIAVTEDIHMTLSRLSGIPVEKLTQADSKKYINLEKEKLRV 518  
 K K S K +KR+ VT+ +A S L+G+P+ +T++S + +NLEKELKRV  
 Sbjct: 463 ISDKIYKSIKEKKRQE--VTDQAVIAVASTLTGVPTTQMTSESRLINLEKELKRV 520  
 55 Query: 519 QDDAVTAISAIRKNQSGIRTKRPIGSPFPLQPTGVGKTELAKALAEVLDDDSALIR 578  
 GQ++A++A+SKAIRK +SG+ +RP+GSFMP+GPTGVGKTELAKALA+ +F E +IR  
 Sbjct: 521 QGEAISAVSKAIRARSGVADSRPMSGPFPLGPTGVGKTELAKALADSVGSEEDMIR 580  
 Query: 579 FDMSEYMEFASHLNGAPPGYVGYDEGGELTEKVRNKPYSVLLPDBVEKHAHPDIFNVLL 638  
 DMSE+MEK + S L GAPPYVGYDEGG+L+TE+VRNKPYSV L DBVEKHA D+FN++L  
 60 Sbjct: 581 VDMSEYMEKHSRLGAPPGYVGYDEGGQLTEKVRNKPYSVLLPDBVEKHALDVPNTML 640  
 Query: 639 QVLDDGV/LDSGRKRVDFSNITIIIMTSMLGATLDDKTVGFGAKDI+SHYFTAMQKRM 698  
 Q+LDDG +TD+G+GRVDF NITIIIMTSMLGATLDDKTVGFGAK+I+ DY+AMQ R+L+E  
 Sbjct: 641 QILDDGV/LDVGKRVDFSNITIIIMTSMLGATLDDKTVGFGAKNITADYAMQSRILE 700

-1630-

- Query: 699 ELKAYRPEFINRIDEKVVHSLSQDNMRVVKIMVKPLILALKDKGDLKQPSALKHL 758  
ELK+ YRPRF+NRIDE+VHSL + +VKIM K LI L ++ +K PSA+K +  
Sbjct: 701 ELKHHYRPEFLARIDENIVHSLSQETBQIVKIMSKSLIKRLAQDHVKLTSPSAIKLI 760
- 15 Query: 759 AEDGYDIRMGARPLRRTIQTVHDLSELLANOVKKGQVIKIGVSKGKLL 809  
AE G+D R GARPLR+ +Q +VED LSR LL+ +K G I IG S K+K  
Sbjct: 761 AEWGDFEYGARPLRKLQKVEDLLSQLLSGEIKAGNHISIGASNKIK 811
- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4537> which encodes the amino acid sequence <SEQ ID 4538>. Analysis of this protein sequence reveals the following:
- Possible site: 44
- >>> Seems to have no N-terminal signal sequence
- 15 INTEGRAL Likelihood = -1.75 Transmembrane 32 - 48 ( 32 - 48)
- Final Results -----
- bacterial membrane --- Certainty=0.1702 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
20 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- RGD motif: 285-287

An alignment of the GAS and GBS proteins is shown below.

- 25 Identities = 618/814 (75%), Positives = 716/814 (87%), Gaps = 1/814 (0%)
- Query: 1 MSHYSIKLQEVRLAQFQAARYBESHVLESWHLLAMVLVHDSVAGLTPAEYBSEVAIEEY 60  
M YS K+Q++FR AQFQAAR++SH LE+WH+LLAMV V +S+A + +HY++VAIEEY  
30 Sbjct: 1 MNHSTKMQDIFRQAQFQAARFDHSHCLCTHWHVLLAMVAVDNSLANMLSEIYDAQVAIEEY 60
- Query: 61 EAATILALGRAPKEITTYQFLEQSPALKKILKLAENISIVVGAEDVGTEHVLAMLVNK 120  
EAA IIA+G+ PKK+++ F QS L +L A+ IS + +VG+EHVL A+L+N  
Sbjct: 61 EAAATILAMGKTKBQLSRVDFRQSKTTLTNLAFQAQAISQITRDQVGSSEHVLFAILLNP 120
- 35 Query: 121 DLAITRILSLVGRFGDQGESV-RMVDLRKALERHAGFTKDDIKAIYELRNHPQKASGAS 179  
D++A+R+LE+ G++ +D+G R+ DLKKA+ERHAG++K+ IKAI+ELR PKK K+ +  
Sbjct: 121 DIMASRLLEIAGYQIKDNGSQPRLADLRKAIERHNGYSKEMIKAIHELRPKPKTKTQGT 180
- 40 Query: 180 FSDMKKPPSTAGLADFTRLDSQMAVDGEIEFVIGRDKEISRMVQLSRKTKNNPVLVG 239  
FSDMKKPPSTAGL+L+DFTRL+L+MA G +E VIGRD+R+SRM+QVLSRKTKNNPVLVG  
Sbjct: 181 FSDMKKPPSTAGLSDFTRLDTEMARQGLLESVIGRDQEVSRMIQVLSRKTKNNPVLVG 240
- 45 Query: 240 AGVGTKALAYGLAQRIANGNI PYELRMRVLELDMMSSVAGTRFRGDFEERMEQIADIE 299  
AGVGTKALAYGLAQRIANG IPYEL++MRVLELDMMSSVAGTRFRGDFEERMEQII DIE  
Sbjct: 241 AGVGTKALAYGLAQRIANGAI PYELKEMRVLELDMMSSVAGTRFRGDFEERMEQIIDIE 300
- 50 Query: 300 EDGHIILFDLHTTMSGGGIDSTLDAANILKPAALRGTLRTVGATTQREYQKHIEKDA 359  
DG IILF+DELHTTMSGGGIDSTLDAANILKPAAL+RGTL VGATTQREYQKHIEKDA  
Sbjct: 301 ADGQIILFVDELHTTMSGGGIDSTLDAANILKPAALRGTLRTVGATTQREYQKHIEKDA 360
- 55 Query: 360 ALGRRFAKLVLEENLEDAYEILLGLKPAYEAPHNVTISDEAVMTAVKVAHRYLTSKNLP 419  
ALGRRFAK+L+EEP EDAY+IL+GLK +YE +HN+IS+EAV TAVK+AHRYLTSKNLP  
Sbjct: 361 ALGRRFAKILIEEPNTEAYQIMSLKLSYETVHNVSINSNAKTVAMHRYLTSKNLP 420
- 60 Query: 420 DSAIDLLDEASATVQNMKKNAPELLTSDVQAILDDDMKSASKALIASYKGGKRPPIAVT 479  
DSAIDLLDEASA VQ M+KK+AP LT +DQA+++ DMK S+ L KG+ RKP VT  
Sbjct: 421 DSAIDLLDEASAAVQNMVKSAPETLTTPIDQALINGDMKKVSRLLAKEAGQMKRPPIVPT 480
- 65 Query: 480 EDHIMATLSRLSGIPVEKLTQADSKKYLNEKEKLHRRVIGQDAVTAISRIRFRNQSGIR 539  
ED I+ATLS+LSGIP+EKLTQADSKKYLNEKEKLHRRVIGQDAVTAISRIRFRNQSGIR  
Sbjct: 481 EDDITATLSKLSGIPLEKLTQADSKKYLNEKEKLHRRVIGQDAVTAISRIRFRNQSGIR 540
- Query: 540 TGKRPIGSFMTLPGTGVGKTELAKALAEVLFDDEALIRFDMSEYMEKFAASHLNGAPPG 599  
TGKRPIGSFMTLPGTGVGKTELAKALAEVLFDDEALIRFDMSEYMEKFAAS LNGAPPG  
Sbjct: 541 TGKRPIGSFMTLPGTGVGKTELAKALAEVLFDDEAALIRFDMSEYMEKFAASHLNGAPPG 600



Query: 600 YVYDGGEGGELTKVRNKKPYSVILFDEVEKAHPDIPNVLLQVLDGILTSRGRKVDPSNF 659  
 YVYDGGEGGELTKVRNKKPYSVILFDEVEKAHPDIPNVLLQVLDGILTSRGRKVDPSNF  
 Sbjct: 601 YVYDGGEGGELTKVRNKKPYSVILFDEVEKAHPDIPNVLLQVLDGILTSRGRKVDPSNF 660

Query: 660 IITMISNGATGALADKDTGVFGAKDISHDYTAQKRIMBELKKAYRPEFNIDEKRVVHF 719  
 IITMISNGATGALADKDTGVFG KDI +A+MRRL+SELK+ YRPFENIDEKRVVHF  
 Sbjct: 661 IITMISNGATGALADKDTGVFGKVIDHQHQAHPERRLSELK+YRPFENIDEKRVVHF 720

Query: 720 SLSDQNMREVVIMKPELILALDKGMDLKPPASALKHLEMDGYDIEMGARPLRTITQ 779  
 SL+QNMRE+VYVIM+PVL I +KG+ LK QF ALKHL+E GYD MGARPLRT+QT+  
 Sbjct: 721 SLTDQNMREVVIMKPELITLAEKGKIDITLAEKGLKIQPLAKLHSEVGYDERHGARPLRTLTQE 780

Query: 780 VEDHLELLLANQVKEGQVIKIGSKGLKFDIA 813  
 +ED LSEL+L+ ++ G +KIG+S GKL F+  
 Sbjct: 781 IEDKLELLLSRELTSHTYIKGLGSHGKFTPIA 814

A related GBS gene <SEQ ID 8819> and protein <SEQ ID 8820> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 9
McG: Discrim Score: -13.52
GvH: Signal Score (-7.5): -2.1
    Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -2.34 threshold: 0.0
    INTEGRAL Likelihood = -2.34 Transmembrane 32 - 48 ( 32 - 49)
    MERIPHERAL Likelihood = 0.95 112
    modified ALOM score: 0.97

*** Reasoning Step: 3

----- Final Results -----
    bacterial membrane --- Certainty=0.1935(Affirmative) < success
    bacterial outside --- Certainty=0.0000(Not Clear) < success
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

The protein has homology with the following sequences in the databases:

```

47.4/69.6% over 804aa
EGAD|136761| ClpC ATPase Insert characterized
GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase Insert characterized

ORF00207 (298 - 2727 of 3045)
EGAD|136761|145854 (2 - 806 of 825) ClpC ATPase {Listeria monocytogenes}
GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase {Listeria monocytogenes}
%Match = 33.6
%Identity = 47.4 %Similarity = 69.6
Matches = 372 Mismatches = 229 Conservative Sub.s = 174

87      117      147      177      207      237      267      297
SFF**SPIIKWYVINDWRAYQ*TSF**FDSIIIR*KNYKT*RKFDSDGDIR**RIERASLCY*SSYAP*IIITIR*KRIP

M

327      357      387      417      447      477      507      537
FMSHYSIKLGVFLAQFOAARYESHYSILYSWILLIAMVLVHDSVAGLITFAEYSEVALESYAPATILAGRAPKEEITNY
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
MGGRFTQRAQVKVLASQEFAMRINHSLIGTETHILLGLVRESEGTAA-KALYEIGLISSEKVKQVEVLIGIG-ERAVTTI

20      30      40      50      60      70

567      597      627      657      687      717      744      774
QLFQPSGLKGVFLKARNISIVVGAEDVYTHFVLLAMVKNKDLATRIKLXVGFRQGDGSESV-RNVDLRKLERHGET
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QTT--PRACKVLISMDPEAKLICTHYVGTETHILGLIGRESEVAVRNLNIGLISINKARQVQVILGLGGDA-----

90      100      110      120      130      140

```



-1633-

**Example 1477**

A DNA sequence (GBSx1563) was identified in *S.agalactiae* <SEQ ID 4539> which encodes the amino acid sequence <SEQ ID 4540>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4541> which encodes the amino acid sequence <SEQ ID 4542>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 179/213 (83%), Positives = 199/213 (92%)

Query: 1  MLIVLAGTIGAGKSSLAALGQHGLGTDVIFYEAVDNNFVLIDYYQDPQKYAFLQLIIFLNK 60
      MLIVLAGTIGAGKSSLAALG+HLGTDVIFYEAVDNNFVLIDYYQDP+KYAFLQLI+FLNK
Sbjct: 1  MLIVLAGTIGAGKSSLAALGQHGLGTDVIFYEAVDNNFVLIDYYQDPKKYAFLLQIYFLNK 60

Query: 61  RFQSIKEAYKANNVLDRSIFEDELFLTYNKGNNVITKELDIYKELLANMLEELEGMPK 120
      RF+SIKEAY+A+NN+LDRSIFEDELFL YNKGNNVITKELDIY+ELLANMLEELEGMPK
Sbjct: 61  RFQSIKEAYQADNNILDRSIFEDELFLKLYNKGNNVITKELDIYQELLANMLEELEGMPK 120

Query: 121 KRPDLLVYIDVSFDKMLERIKRGRSFEQVDSNPFLDYDYKQVHSEYPEWYENYDVSPKI 180
      KRPDLL+YIDVSFDKMLERI++RGRSFEQVD NP L YY QVH EYP WYE-Y+VSPK+
Sbjct: 121 KRPDLLIYIDVSFDKMLERIKRGRSFEQVDGNPSLEQYHQVHGEPYTWYEDYVSPKM 180

Query: 181 RIDGNKLDVFVNPEDLQHVLTIDISELQKLDLL 213
      +IDGN LDFV+NP+DL VL ID++L+L LL
Sbjct: 181 KIDGNSLDFVQNPDLATVLKMDITKLKELHLL 213

```

A related GBS gene <SEQ ID 8821> and protein <SEQ ID 8822> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 0
McG: Discrim Score: 3.94
GvH: Signal Score (-7.5): 1.42
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 7.69 threshold: 0.0
PERIPHERAL Likelihood = 7.69 49
modified ALOM score: -2.04

*** Reasoning Step: 3

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

-1634-

SEQ ID 4540 (GBS9) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 5; MW 52kDa) and Figure 12 (lane 2 & 3; MW 50.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 6; MW 27kDa) and Figure 3 (lane 2; MW 25kDa). The GBS9-GST fusion product was purified (Figure 191, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 318), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1478

- 10 A DNA sequence (GBSx1564) was identified in *S.galactiae* <SEQ ID 4543> which encodes the amino acid sequence <SEQ ID 4544>. Analysis of this protein sequence reveals the following:

Possible site: 36  
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1182 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4545> which encodes the amino acid sequence <SEQ ID 4546>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 30 An alignment of the GAS and GBS proteins is shown below.

Identities = 281/323 (86%), Positives = 305/323 (93%)

- Query: 3 QLNSSFMIGKVEI PHRTVLAPMAGITNSAPRTIAKEFGAGLVVMMISEKGLLYNNEKTL 62  
+LNSFF IG VEI PHRTVLAPMAG+TNSAPRTIAKEFGAGLVVMMISEKGLLYNNEKTL  
35 Sbjct: 27 KLNSSFRIGDVEI PHRTVLAPMAGVITNSAPRTIAKEFGAGLVVMMISEKGLLYNNEKTL 86
- Query: 63 HMLHIDENHFMISIQLFGGDABGLKRAADFIQSNFKADIVDINMGCPVNVKVGNEAGAKW 122  
HMLHIDENHFMISIQLFGGDABGLKRAADFIQ+HTKADIVDINMGCPVNVKVGNEAGAKW  
Sbjct: 87 HMLHIDENHFMISIQLFGGDABGLKRAADFIQNTKADIVDINMGCPVNVKVGNEAGAKW 146
- 40 Query: 123 LRDPEKIYHIVKIVTSVLDIPLTVIOMRTGWSDSSNAENALAESAGVSALAMHGRTREQ 182  
LRDP+KIYHIVKIVTSVLDIPLTVIOMRTGW+DSS A+ENALAESAGVSALAMHGRTREQ  
Sbjct: 147 LRDPEKIYHIVKIVTSVLDIPLTVIOMRTGWSDSSLAENALAESAGVSALAMHGRTREQ 206
- 45 Query: 183 MYTGCDHETLGVKAVTSIPFIANGDIRTVHDAKMIEBIGADAIMVGRGARSNPYIF 242  
MYTGCDHETL +V+KA+T IFFI NGD+R+V DAKFMIEBIG DA+M+GR A +NPY+F  
Sbjct: 207 MYTGHDETHARVSKAITKIPFIANGDVRVQDAK+MIEBIGDAVMIGRAAMNPYIF 266
- 50 Query: 243 TQINHHFFETQELPDLFFERMLDVAEDHITRLVNLKGETIAVREFRGLAPHYLGRKSGAA 302  
TQINHHFFETG+ LPDLPF K ID+A+DHL RI+NLKGETIAVREFRGLAPHYLGRG +GAA  
Sbjct: 267 TQINHHFFETQELPDLFFAKKILDIADHLKRLINLKGETIAVREFRGLAPHYLGRGTAGAA 326
- Query: 303 KIRGAVSRAETLAHVQELFAGLR 325  
K+RGAVSRAETLAHV+ +F +R  
55 Sbjct: 327 KVRGAVSRAETLAHVRAIFETVR 349

-1635-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1479**

A DNA sequence (GBSx1565) was identified in *S. agalactiae* <SEQ ID 4547> which encodes the amino acid sequence <SEQ ID 4548>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2164(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 3930:

```
Identities = 235/288 (81%), Positives = 259/288 (89%)

Query: 1 MDKIIKSISTSGSFAYVLDCTETVTAQEKHQTLSSTVALGRTLIANQILAAQKQNS 60
      MDKIIKSI+ SG+FRAYVLD TETV AQEKH TLSSTVALGRTLIANQILAAQKQ+S
Sbjct: 1 MDKIIKSIAQSGAFRAYVLDSTETVALAQEKHNTLSSTVALGRTLIANQILAAQKQNS 60

Query: 61 KVTVKVIGDSSFGHILSVADTNGVKGVIQMTGVDIKKTATGEVLVGFPMNGHFPVITD 120
      K+TVKVIGDSSFGHILSVADTNG+VKGVIQMTGVDIKKTATGEVLVGFPMNGHFPV I D
Sbjct: 61 KITVKVIGDSSFGHILSVADTNGVKGVIQMTGVDIKKTATGEVLVGFPMNGHFPVITD 120

Query: 121 YATGQPTTSTPLITGEIGEDFAYVLTESECTPSAIGLVNVLDDDKVKVAGGFMVLQVL 180
      Y TG PYTSTPLITGEIGEDFAYVLTESECTPSA+GLNVLLD+ DIVKVAGGFM+QVL
Sbjct: 121 YGTGNPTTSTPLITGEIGEDFAYVLTESECTPSAIGLVNVLDDDKVKVAGGFMVQVL 180

Query: 181 GASDEEISRYEKRIGQEMPSISLSENNHIESILSAIYGEDDYKRLSEDSLAFTDCDSKE 240
      GAS+EEI+RYEKR+QEMF+IS LL S+NH++LL AIYG+ YKRLE+ L+F CDCS+E
Sbjct: 181 GASEEETARYEKRIGQEMPAISHLASKNHVDALLAITYGDEPKRLSEELSLFQCCDSKE 240

Query: 241 RFEALMTLSTKEIQAQMKDEKGVITCOFCKQTYPTFEDELEKIND 288
      RFEAL+TL +LQAM DEDKG ET CQFC Y P E DLE II+D
Sbjct: 241 RFEALMTLSTADLQAMIDEDKGAIVCOFGTKYQFNESGLEAISD 288
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1480**

A DNA sequence (GBSx1566) was identified in *S. agalactiae* <SEQ ID 4549> which encodes the amino acid sequence <SEQ ID 4550>. This protein is predicted to be surface-located membrane protein 1 (Imp1). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4312(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AA929480 GBS:AF019377 tellurite resistance protein [Rhodobacter
      sphaeroides]
Identities = 64/350 (18%), Positives = 146/350 (41%), Gaps = 7/350 (2%)
```

-1636-

- Query: 44 LTPAQKSAISEKTPALVDITPVQDQNALDPQGSABGVNTVNHILSEQKKIQIPQVDEL 103  
 L A E + + V D + + FG A + T + L + K + D  
 Sbjct: 34 LASAPPEKAQETIRRMARLNVSDSQSIIGPGSKAQKILQITISQQLADLVNKNKDVGPAGDS 93
- 5 Query: 104 LKNANRELNGFIKDYKDTAPAELEKPKNLQKLFQSKTSIQEFPYFDSQNTREQKMDMAA 163  
 L + + GF + + + E + + L + + F + + + Q + D +  
 Sbjct: 94 LREVVSTIRGF ---- SVSEFVRRKASWERRLLGRT - APPARFVARYEDVQQQIDRITQ 147
- 10 Query: 164 NVVQSDITLARNIVSAEMLIRNTKSIKHLGVIAFTSSQAEANRASHLQOEITLALDS 223  
 + + E L + I + L + L IA + A R + + +  
 Sbjct: 148 SLLTHSRLLKIKIGLDILYARTLDYDELALYIAGDEVADLDRVIPAKEAEVAATP 207
- Query: 224 QTSEYQIKSNQLARMTVEVINTLEQHPHYVSRLYVAMATTPQMRNLVKVSSIMRQKLGML 283  
 + IK + L + + LE + + V + P + R + + + +  
 15 Sbjct: 208 E - GDRMIKAEKLRDLRAARDLERRVHDLKLTQVIMQSLPSIRLVQENDKALVTRINST 266
- Query: 284 RNTATPIMKLSIAQLGMMQGSVKSGVTADAINVANNALQMLAETSKEAIPMLEKTAQSP 343  
 NT + P + + HQ + Q + S + + N L AE + + K +  
 Sbjct: 267 LVNTVPLMETQLAQAVITQRSKRAAKAVRGASDLTNELLTANAEMLQKANKIVRKEMERG 326
- 20 Query: 344 TVSIKSVTALAESLVAQNGIILAAIDHGRKERAQLESIAVSIKSAETINDSV 393  
 I + V + L + A N + A D + GR RA E + + + D + +  
 Sbjct: 327 VFDIEAVKKNATLITATINESLAIADGRARRATATETLQMEAEALADTL 376
- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4551> which encodes the amino acid sequence <SEQ ID 4552>. Analysis of this protein sequence reveals the following:
- Possible site: 55  
 >>> Seems to have no N-terminal signal sequence
- 30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3230 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- 35 An alignment of the GAS and GBS proteins is shown below.  
 Identities = 333/413 (80%), Positives = 379/413 (91%)
- Query: 5 FNFIDQIADNAITKTDKTEIISNQTSIQGIAFFPEKLTFRQKSAISEKTPALVDITFV 64  
 FNFIDQIADNAI KTDKTTIIS+ T GQI+FYEKL+ Q++IA+ K PALVDITF+  
 40 Sbjct: 4 FNFIDQIADNAVIKTDKTDIISDLPTITNGQISFFPEKLGADQQTAITAKAPALVDITFL 63
- Query: 65 DQNALDPQGSABGVNTVNHILSEQKKIQIPQVDELHNNRELNGFIKDYKDTAP 124  
 DQNALDPQGSABGVN TVNHIL+EQKK+IQPVDDILK+ NRELNGFIKDYKDTAP  
 Sbjct: 64 ADQNALDPQGSABGVNATVNHILAEQKKLQIPQVDDILKSNRELNGFIKDYKDTAPV 123
- 45 Query: 125 ELEKPKNLQKLFQSKTSIQEFPYFDSQNTREQKMDMAANVVKQSDITLARNIVSAEMLIE 184  
 + L KKN + QKLFQKS + LQEFYFDSQNTREQKMD MA VVKQSDITLARNIVSAEMLIE  
 Sbjct: 124 DLDKKNPFLQKLFQSKSDITQEFYFDSQNTREQKMDMAANVVKQSDITLARNIVSAEMLIE 183
- 50 Query: 185 DNTKSIENHGVIAFTSSQAEANRASHLQOEITLALDSQTSEYQIKSNQLARMTVEVINT 244  
 DNTKSIH+LGVIAFT+SQ EA+ RA+ LQ++ DS T+YQIK+ LAR TEVINT  
 Sbjct: 184 DNTKSIENHGVIAFTSSQAEARAAALQKDLTKDSATPDYQIKADLLARTTEVINT 243
- 55 Query: 245 LEQHPHYVSRLYVAMATTPQMRNLVKVSSIMRQKLGMLRNTTIPIMKLSIAQLGMMQGS 304  
 LEQHP HY+SRLYVAMATTPQMRNLVKVSSIMRQKLGMLRNTTIPIMKLSIAQLGMMQGS  
 Sbjct: 244 LEQHPHYVSRLYVAMATTPQMRNLVKVSSIMRQKLGMLRNTTIPIMKLSIAQLGMMQGS 303
- Query: 305 VKSGVTADAINVANNALQMLAETSKEAIPMLEKTAQSP TVSIKSVTALAESLVAQNGI 364  
 VKSG+TADAI+NANNAALQMLAETSKEAIP LE+AQ+PT+S+KGV+LAESLVAQNGI  
 60 Sbjct: 304 VKSGMTADAINNANNAALQMLAETSKEAIPALQSAQNPTLSMKGVTSIASELVAQNGI 363
- Query: 365 IAAIDHGRKERAQLESIAVSIKSAETINDSVKIKDKKIVKALLNKGSTORKVDV 417  
 IAAID GRKERAQLESA+I+SAETINDSVK+KD+ IV+ALL+EGK TQ+ +D+  
 Sbjct: 364 IAAIDHGRKERAQLESIAITSAETINDSVKLRDQNTVQALLSEKGETQKTIDK 416
- 65

-1637-

SEQ ID 4550 (GBS201) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 5; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 3; MW 74.5kDa) and in Figure 62 (lane 8 & 9; MW 74.5kDa). The GBS201-GST fusion product was purified (Figure 209, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 304), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1481

- 10 A DNA sequence (GBSx1567) was identified in *S.agalactiae* <SEQ ID 4553> which encodes the amino acid sequence <SEQ ID 4554>. This protein is predicted to be rhotry protein. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have an uncleavable N-term signal seq  
 15 INTEGRAL Likelihood = -6.58 Transmembrane 13 - 29 ( 10 - 31)  
 INTEGRAL Likelihood = -1.54 Transmembrane 33 - 49 ( 33 - 49)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3633 (Affirmative) < succ>  
 20 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4555> which encodes the amino acid sequence <SEQ ID 4556>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have a cleavable N-term signal seq.  
 30 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 35 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 115/239 (48%), Positives = 162/239 (67%), Gaps = 3/239 (1%)  
 40 Query: 32 EVIATLLIGSGYCYVYVD-KKRLKRFSTNQRTEALKSDIKETDQIRHLRLKKDNRS 90  
 +++ + I G GY + V +KRL + +++E LK+ I+ D+ +R L+ D+  
 Sbjct: 42 DILPAIAIGTGVALFRVGHGKRLAKAKIAKQLEDLQKIQADLRKVLDTYLAHDH 101  
 Query: 91 KEVIELAHQITLPOLDLIRNANQLQKAIERNIYKRITKKANTFSNEINQIKLHASP 150  
 +Y LA QHLPOL I+ +A L+ ++P IY+RITKKAN ++I QL L + L  
 45 Sbjct: 102 PQYNVLAQQLLPQLSDIKAKAITLKDQLDPQIYRITKKANDVESDITLQETLQIATTL 161  
 Query: 151 --EPISDQDEEMIRIAPLEKPPYTHNQDDHFAITLKRTKREANKEALAJAHQANMKRFTDV 208  
 +P+ +I APLEKP+Y NIQ DH AIL KI+ ADN+ EL A+H ANH+RF D+  
 50 Sbjct: 162 NPQFLKTPSPNLNNKAPLEKPPYTHNQDDHFAITLKRTKREANKEALAJAHQANMKRFTDV 221  
 Query: 209 LAGYIRIKQSPKPNNAKERLEQALQATKKPRLDLDETLLRQLNESMKKDFVSLRNMQG 267  
 L GY++IK+ PKN+ NA RLEQA QAI++F+ DIAETLR+LINESD+KDFD+SLR+MQG  
 Sbjct: 222 LAGYLIKKEEPKNYNAARLEQAQATQDFDELDERTLRINESDLKDFDLSLRNQG 280

SEQ ID 4554 (GBS265) was expressed in *E. coli* as GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 2; MW 56kDa) and in Figure 62 (lane 6; MW 56.3kDa).

The GBS265-GST fusion product was purified (Figure 207, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 258A) and FACS (Figure 258B). These tests confirm that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1482

A DNA sequence (GBSx1568) was identified in *S. agalactiae* <SEQ ID 4557> which encodes the amino acid sequence <SEQ ID 4558>. This protein is predicted to be glutamate-cysteine ligase (gshA). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -1.70    Transmembrane    575 - 591 ( 575 - 591)

----- Final Results -----
          bacterial membrane --- Certainty=0.1680 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AA085808 GB:AB004933 glutamate--cysteine ligase [Pseudomonas aeruginosa]
Identities = 142/468 (30%), Positives = 220/468 (46%), Gaps = 62/468 (13%)
```

```
Query: 12  SHLPIL-QATPGLERESLRHQPTQVQAQTPHPKTLGSRNHYPIQCTYDEPQELLETPI 70
      ++L+L + G+ERE LR+ ++A TPHP+ LGS HP I TDYSE LE ITP
Sbjct: 16  ANLPLLTBCLHGIERECLRVSDG-KIALTTPHRLGSLTTHPQITTYSEALLETPTT 74

Query: 71  AKDSQRAIRFLKATSDVAGRSINHDEYLNPLSMPPKV-REEDIQIAQLEDA---PEYDY 125
      D ++ L+ I A ++ EYLN SMP ++ EE I IA+ + ++Y Y
Sbjct: 75  ETDVADTLGLDERHRFASKLD-GRYLNPSMPCLEPDESIPIARYGSSMIGRLKVVY 133

Query: 126 RKLYEKTYGKLIQISIGIHYNLGLQELLTSLFELSQAD-NAIDFQNLNKLKSNFLRY 184
      RK L YGK +Q I+GIHYN L + L L + ++ + D+Q+ Y+ L +HF RY
Sbjct: 134 RKGLALRYGKWTQCIAGIHYNPLSLPERLNPLLRQAEGSELSERDYQSPAYIALIRIFRY 193

Query: 185 RWLLTYLYGASPVAEEDFLDQKLNPNVR-----SLRNSHLGYVNHKDIRIS-- 230
      WLL YL+GASP + FL + + R SLR S LGY N+ ++
Sbjct: 194 SWLLMYLYGASPALDAGSLRGRPSQLERLDEHTLYLPATSLRMSDLGYNNAQAGLTPC 253

Query: 231 YTSLKDYVNDLENV-----KSQQLIARKEFYSPVRLR-----G 264
      Y L+ Y+ L AV + L E E+YS +R + G
Sbjct: 254 YNDLQSYIDSLRQVSTPYTPPYEKVGTGKODGSEWGLATNMLQIRNEYYSIRPKRVYITG 313

Query: 265 SKACRNYLEKGITYLEFRTFDLNPPSPIGITQETVUTVHLFLALLWIDS----- 314
      + + +G+Y+R R D+NPF P+GI + + FLL + DS
Sbjct: 314 ERPFQALAAAGVQYEVVRCLDINPLPLGLDLDKARFLDAPLFLCPAFSDSPLINGGCSDA 373

Query: 315 SSHIDQDIKEANKRN-DLIALSHPLEKLPNQAPVSDLDVAMQSVIHPNLSPTYQDLES 373
      ++ + +KE R L P+R + + + + + + L +
Sbjct: 374 TDNFLAVVKGRRPGLQIQRRCQFVELQVWANELLERIADTAALLDRARGGEFAHAALAA 433

Query: 374 VKRQIQSPELTVAQQLLSMI--EGLSLETFGQRQQIYHYDAWEAPYA 419
      + ++ ELT + Q+L+++ G S E F RQ + + +Y + P A
Sbjct: 434 QRAKVADELTPSAQVLKVRIRKRCSEAFSLAQGRHAYYFRQHPLA 481
```



-1639-

There is also homology to SEQ ID 4560.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1483

- 5 A DNA sequence (GBSx1569) was identified in *S. agalactiae* <SEQ ID 4561> which encodes the amino acid sequence <SEQ ID 4562>. Analysis of this protein sequence reveals the following:

Possible site: 59  
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1504 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB73814 GB: AL139078 helix-turn-helix containing protein  
[Campylobacter jejuni]  
Identities = 107/223 (47%), Positives = 148/223 (65%), Gaps = 7/223 (3%)

- 20 Query: 1 MDKEKLDYWKTIITFLRNVLGDNYEIVLHVVDENDIYIGELVNSHISGRISSEPLTTFAL 60  
MD+ + + + FL VLG+ YEIV HV+ E+ YI + NSHISGR++ SPLT FA  
Sbjct: 1 MDEGQKQCFIKLITYPLGEVLGRQYEIVFHVITEDGAYIAAIAHSHISGRSLDSPITAFAS 60
- 25 Query: 61 DLIKNKVKYKEDFVINYKAIVSPLNKEVRGSTFFIKNAQNELEGMCLNLDISAYQNIAL 120  
+L+ NK Y EKD+ + YKA+ V + K + RGSTFFIKN ++L G+LCIN D S +++  
Sbjct: 61 ELMQNKYKLEKDFLCIDYKALVGR-SKLIRGSTFFIKN-HDKLVGILCINHDSITMRDLIC 118
- Query: 121 DILDVLNVL-NVNKILPKSPQKISLPQQEFPVLSGNIQDIISIVDPSLLNQNHLSQ 179  
++DL + ++ IL IS Q + + E LS +I+DI+ + VD S LN+ LS  
30 Sbjct: 119 KMIDLEKIGDMGDLIL---GNISPSQNDSSIETLGHSTEDILVQSVDSYLSNDYQESIT 174
- Query: 180 VKVBIYSKLEHKGVPQLKGAVSKVAEVLNISEPSVYRYLKKIE 222  
K EI KL+ERG+P +KGAV VA+ L ISEPSVYRYLKK +  
35 Sbjct: 175 QKEELAEKLYKGIPTNKGAVPIVAKFLKISEPSVYRYLKKFK 217

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4563> which encodes the amino acid sequence <SEQ ID 4564>. Analysis of this protein sequence reveals the following:

Possible site: 21  
>>> Seems to have no N-terminal signal sequence

- 40 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1636 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 169/224 (75%), Positives = 198/224 (87%), Gaps = 3/224 (1%)

- 50 Query: 1 MDKEKLDYWKTIITFLRNVLGDNYEIVLHVVDENDIYIGELVNSHISGRISSEPLTTFAL 60  
MDKE L+YWKTI+ITFLH+VLGDNYRI+LHV+D+NDIYIGELVNSHISGR+ SPLTTFAL  
Sbjct: 1 MDKETLNYWKIVITFLHVLGDNYRIILHV+DNDIYIGELVNSHISGRSKQSEPLTTFAL 60
- Query: 61 DLIKNKVKYKEDFVINYKAIVSPLNKEVRGSTFFIKNAQNELEGMCLNLDISAYQNIAL 120  
DLI NKVKYKEDFVINYKAIVSP+KEVRGSTFFIK+ + LEGLCLNLDISAYQ +A  
55 Sbjct: 61 DLTINKVKYKEDFVINYKAIVSPQEKEVRGSTFFIKDKKGNLEGMCLNLDISAYQGVAR 120
- Query: 121 DILDVLNVLNVNKKILP--KSPQKISLPQQEFPVLSGNIQDIISIVDPSLLNQNHLSQ 178  
D+L LVNIN+ +P K P+ ++ PQ EE VE+L+ NIQDII +I+DPSLL N+HLSQ

-1640-

Sbjct: 121 DLKLVNINLEHFIPTAKPKTIVT-PQPREAVILTSNIQDIIGQIIDPSLLRHVLSQ 179

Query: 179 EVKKEIVSKLHEGVPQLKGNVSKVAEVLNISEPSVRYLKKIE 222  
+VK++IV+KL+KEGVFQLKGNVSKVA++L ISEPSVRYLKKIE

5 Sbjct: 180 DVKIDTAVKLYEKGVPQLKGNVSKVADILCISEPSVRYLKKIE 223

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1484

- 10 A DNA sequence (GBSx1570) was identified in *S. agalactiae* <SEQ ID 4565> which encodes the amino acid sequence <SEQ ID 4566>. This protein is predicted to be regulatory protein pfoR. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

- 15 INTEGRAL Likelihood = -7.80 Transmembrane 299 - 315 ( 296 - 325)  
INTEGRAL Likelihood = -7.54 Transmembrane 172 - 188 ( 169 - 193)  
INTEGRAL Likelihood = -7.17 Transmembrane 71 - 87 ( 66 - 98)  
INTEGRAL Likelihood = -4.99 Transmembrane 261 - 277 ( 260 - 278)  
20 INTEGRAL Likelihood = -2.81 Transmembrane 128 - 144 ( 127 - 149)  
INTEGRAL Likelihood = -2.18 Transmembrane 101 - 117 ( 101 - 119)  
INTEGRAL Likelihood = -0.53 Transmembrane 198 - 214 ( 197 - 214)

----- Final Results -----

- 25 bacterial membrane --- Certainty=0.4121 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA60239 GB:X86525 pfoS [Clostridium perfringens]  
Identities = 96/147 (65%), Positives = 122/147 (82%)

- 30 Query: 100 GNGIIPGFIAGYLVGSPVFKMERNIPQGLDLISIIIGAPLTRLVAKLLTFLINSTITLT 159  
G GI+PGF+AGYL F+K++E+ ID GLDLI II++GAPL R +A + PL+ +TL I  
Sbjct: 1 GFGILPGFIAGYLVGSPVFKLEKKIPAGLDLIVIVLGAFLVRGIAAISHPLVETTLQNI 60

- 35 Query: 160 GDILTSQAHSNPILMGIIIGTIVVATAPLSSMALTAMGLTGMFMAIGALSVFGSSFM 219  
G ++T+ ++PT+NGIILGG + VVATAPLSSMALTAMGLTGMFMAIGAL+VFGSSFM  
Sbjct: 61 GGVITATSTASPIIMGIIIGTIVVATAPLSSMALTAMGLTGMFMAIGALVFGSSFM 120

- 40 Query: 220 NGVLPFKLLGSRKDNIAFAVEPLTQA 246  
N V F K+K GS+ND IA A+EPLTQA  
Sbjct: 121 NLVFPGRMKPGSKDITLAVIEPLTQA 147

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4567> which encodes the amino acid sequence <SEQ ID 4568>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

- 50 INTEGRAL Likelihood = -8.70 Transmembrane 303 - 319 ( 296 - 325)  
INTEGRAL Likelihood = -7.11 Transmembrane 70 - 86 ( 66 - 98)  
INTEGRAL Likelihood = -6.53 Transmembrane 172 - 188 ( 169 - 193)  
INTEGRAL Likelihood = -4.83 Transmembrane 261 - 277 ( 260 - 278)  
INTEGRAL Likelihood = -2.55 Transmembrane 101 - 117 ( 101 - 119)  
INTEGRAL Likelihood = -2.28 Transmembrane 124 - 140 ( 124 - 140)  
INTEGRAL Likelihood = -1.91 Transmembrane 198 - 214 ( 197 - 215)

----- Final Results -----

- 55 bacterial membrane --- Certainty=0.4482 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

```
>GP:CAA60239 GB:X86525 pfoS [Clostridium perfringens]
Identities = 95/147 (64%), Positives = 123/147 (83%)

5   Query: 100 GTGIIPGVAGYVVSFLIKWNEKNIPGGLDLISIIIVGAPLRLPLAQLITPPVINSTLLTI 159
      G GI+PGF+AGY+ SP+IK++EK IP GLDLI I++GAPL R +A + P++ +TL I
      Sbjct: 1   GFGILGPGFIAGYLGSPVFLKLEKKIPAGLDLIVIVLGAFLNVRGINAISNPLVETTLQNT 60

10  Query: 160 GDILTSSANSNPIIMGMILGGTIVVATAPLSSMALTAMGLGTGIPMAIGALSVFGSSFM 219
      G ++T+++ ++PI+MG+ILGG + VVATAPLSSMALTAMGLGTG+PMAIGAL+VFGSSFM
      Sbjct: 61  GGVITATSTASPIIMGHIILGGIVTVVATAPLSSMALTAMGLGTGPMAGALAVFGSSFM 120

15  Query: 220 NGVLFYRLKLGKRDNIAPAIPLTQA 246
      N V F ++K G +KD IA AIEPLTQA
      Sbjct: 121 NLVFFGKMGKPGSKKDTIAVAIEPLTQA 147
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 302/339 (89%), Positives = 330/339 (97%)

20  Query: 1   MNIIIGTSLILLVLAIFTLFNYPKAPYGTGKMGALASAAACASFLVEAFQDSFFGKVLGQF 60
      M+IIIGTSLILLVLAIF+LFNYPKAP+G KMGALASAAACASFLVEAFQDSFFGKVLGQF
      Sbjct: 1   MDIIIGTSLILLVLAIFSLFNYPKAPHGAKMGALASAAACASFLVEAFQDSFFGKVLGQF 60

25  Query: 61  LSEVGANGSLSGVAAAILVATAIGVTPGYAVLIGLSVSGTGIIIPGLAGYLVFLVKWM 120
      LSEVGANGSLSGVAAAILVATAIGV+PGYAVLIGLSVSOTGIIIPG+AGY+V FL+KWM
      Sbjct: 61  LSEVGANGSLSGVAAAILVATAIGVSPGYAVLIGLSVSGTGIIIPGVAGYVVSFLIWM 120

30  Query: 121 ERNIPGGDLISIIIIIGAPLRLVAKLITPLINSTLLTIGDILTSQAHNSPILMGIILGG 180
      E+NIIPGGDLISIIII+GAPLTR +A+L+TF+INSTLLTIGDILTS A+SNPI+MG+ILGG
      Sbjct: 121 EKNIIPGGDLISIIIVGAPLRLPLAQLITPPVINSTLLTIGDILTSANSNPIIMGMILGG 180

35  Query: 181 TIVVATAPLSSMALTAMGLGTGMPMAIGALSVFGSSPMNGVLFHKLKLGSRKONTAFV 240
      TIVVATAPLSSMALTAMGLGTG+PMAIGALSVFGSSPMNGVLP++LKLK RKNIAFA+
      Sbjct: 181 TIVVATAPLSSMALTAMGLGTGIPMAIGALSVFGSSPMNGVLFYRLKLGKRDNIAPAI 240

40  Query: 241 EPLTQADVTSANPIPIYVTFNVGGAACGLIALMLKLVNDTPGTATPIAGFAVMFAYNPIM 300
      EPLTQADVTSANPIPIYVTFNVGGAAC+LIALMLKLVNDTPGTATPIAGFAVMFAYNP+
      Sbjct: 241 EPLTQADVTSANPIPIYVTFNVGGAACGLIALMLKLVNDTPGTATPIAGFAVMFAYNPVA 300

45  Query: 301 KVLITALGCIILSLLAGYFGGIVPKDYKLVTKKEELQARD 339
      KVLITALGCI++SL+ GY GK VFK+Y+LVTK+ELQAR+
      Sbjct: 301 KVLITALGCIILSLVIGYIGSVFKNRYLVTKCELQARN 339
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1485

A DNA sequence (GBSx1571) was identified in *S. agalactiae* <SEQ ID 4569> which encodes the amino acid sequence <SEQ ID 4570>. This protein is predicted to be adenylosuccinate synthetase (purA). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55  bacterial cytoplasm --- Certainty=0.0560 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

-1642-

>GP:CAB16079 GB:Z99124 adenylosuccinate synthetase [Bacillus subtilis]  
Identities = 320/427 (74%), Positives = 378/427 (87%)

5 Query: 1 MTSVVVGTQWGDGSGKKTIDFLSADAEVIARYQGGDNAGHTTIVIDNKKFKLHLIPSGIF 60  
M+SVVVGTQWGDGSGKKTIDFLSADAEVIARYQGGDNAGHTT D +KLHLIPSGIF  
Sbjct: 1 MSVVVGTQWGDGSGKKTIDFLSADAEVIARYQGGDNAGHTTKFDGTTKHLHLIPSGIF 60

10 Query: 61 FKEKISVIGNGVVNNPKSLVKELAYLHGEGVTTMDLRISDRAHVILPYHKLQDLQEDAK 120  
+K+K VIGNG+VV+FK+LV ELAYLR V+TMDLRIS+RAHVILPYH+KLD++E+K R  
Sbjct: 61 YKDKTCVIGNGVVDDKALVITELAYLHERNVSTMDLRISRAHVILPYHKLDEVEERER 120

15 Query: 121 GGNKIGTTIKGIGPAYMDKAAKRVGIRIADLLDREVFARLKLINLAERNLFEKMYDSTPL 180  
G NKIGTT KGIGPAYMDKAAK+GIRIADLLD+ FAE+L+ NL EKNRL EKM++  
Sbjct: 121 GANKIGTTKKGIGPAYMDKAAKRGIRIADLLDRDAFAELERNLERKNRLLEKMYETSGF 180

20 Query: 181 EFDIDFEEYEEYGGQIKQYVDTDSVILNDALDAGKRVLFESAGQVMIDIQGTYFPVFTSS 240  
+DI +EYEEYGGQIK+YV DTSV+LNDALD G+RVLFESAGQVMIDIQGTYFPVFTSS  
Sbjct: 181 KLEDILDEYEEYGGQIKQYVDTDSVILNDALDGERVLFESAGQVMIDIQGTYFPVFTSS 240

25 Query: 241 NPFVAGVITIGSGVGSFKINKVGVGCKAYTSKRGDGPFPTELFEVGDRIREIGKEGXTT 300  
NPFVAGVITIGSGVGP+KI VVG KAYT+RVGDGPFPTEL DE+GD+IRE+G+EXGTT  
Sbjct: 241 NPFVAGVITIGSGVGPTKIKHVGVGSKAYTTRVGDGPFPTELKDEIGDQIREVGREXGTT 300

30 Query: 301 GRPRRGVWFDSSVVRHSRKRVSITNLSINSDVLSGLDTVKICVAYDLDGKRIDYFPASL 360  
GRPRRGVWFDSSV+RH+RRVSGIT+LNSINSDV+G+T+KICVAY G+I+P+SL  
Sbjct: 301 GRPRRGVWFDSSVVRHARRVSGITDLSINSDVLAGIETLKCICVAYRNGEITIEFPASL 360

35 Query: 361 BQLKCKFIYEELPGNSREDITACRSIDDLFENARNTVRRVSELGVRISTFSGVGRBQT 420  
+L C+P+YEE+KGN+EDIT +SL+LFENAR+Y+RV +L+G +S FSGVP R QT  
Sbjct: 361 KALASCEPVYEEPMGWTEDITGAKSLSELFENARHTLERVSLQGTPLSPFSGVGRBQT 420

Query: 421 NILESVM 427  
N+L SV+  
Sbjct: 421 NVLRSYV 427

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4571> which encodes the amino acid sequence <SEQ ID 4572>. Analysis of this protein sequence reveals the following:

Possible site: 27  
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0560 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 406/430 (94%), Positives = 421/430 (97%)

50 Query: 1 MTSVVVGTQWGDGSGKKTIDFLSADAEVIARYQGGDNAGHTTIVIDNKKFKLHLIPSGIF 60  
MTSVVVGTQWGDGSGKKTIDFLSADAEVIARYQGGDNAGHTTIVID NKKFKLHLIPSGIF  
Sbjct: 1 MTSVVVGTQWGDGSGKKTIDFLSADAEVIARYQGGDNAGHTTIVIDGKKFKLHLIPSGIF 60

55 Query: 61 FKEKISVIGNGVVNNPKSLVKELAYLHGEGVTTMDLRISDRAHVILPYHKLQDLQEDAK 120  
F +KISVIGNGVVNNPKSLVKELAYLH RGVTMDLRISDRAHVILPYH+LDQLQEDAK  
Sbjct: 61 FPKKISVIGNGVVNNPKSLVKELAYLHGEGVTTMDLRISDRAHVILPYHKLQDLQEDAK 120

60 Query: 121 GGNKIGTTIKGIGPAYMDKAAKRVGIRIADLLDREVFARLKLINLAERNLFEKMYDSTPL 180  
GNKIGTTIKGIGPAYMDKAAKRVGIRIADLLD++FAERL+INLAERNLFEKMYDSTPL  
Sbjct: 121 GGNKIGTTIKGIGPAYMDKAAKRVGIRIADLLDKDIFARLKLINLAERNLFEKMYDSTPL 180

65 Query: 181 EFDIDFEEYEEYGGQIKQYVDTDSVILNDALDAGKRVLFESAGQVMIDIQGTYFPVFTSS 240  
+FD IFEYEE YGQ+IKQYVDTDSVILNDALDAGKRVLFESAGQVMIDIQGTYFPVFTSS  
Sbjct: 181 DFDIIFEYEEYAGQIKQYVDTDSVILNDALDAGKRVLFESAGQVMIDIQGTYFPVFTSS 240

Query: 241 NPFVAGVITIGSGVGSFKINKVGVGCKAYTSKRGDGPFPTELFEVGDRIREIGKEGXTT 300

-1643-

NPVAGGVITIGSGVGP+KINKVGVCKAYTSRVGDGPFPTLDFEWG+RIR+G EYGITTT  
 Sbjct: 241 NPVAGGVITIGSGVGP+KINKVGVCKAYTSRVGDGPFPTLDFEWGERIREVGREYGITTT 300

Query: 301 GRPRRVGVFDSVVMRHSRRVSGITNLSLNSIDVLSGLDTVKICVAYDLGKRIDYYFASL 360  
 GRPRRVGVFDSVVMRHSRRVSGITNLSLNSIDVLSGLDTVKICVAYDLGKRIDYYFASL  
 Sbjct: 301 GRPRRVGVFDSVVMRHSRRVSGITNLSLNSIDVLSGLDTVKICVAYDLGKRIDYYFASL 360

Query: 361 EQLKRCCKPIYERLPGWSEDITACRSDDLPLPBNARNYVRRVGLGVGRISTPFSVVGPREQT 420  
 EQLKRCCKPIYERLPGW EDIT RSLD+LPBNARNYVRRVGLGVGRISTPFSVVGPREQT  
 Sbjct: 361 EQLKRCCKPIYERLPGWSEDITACRSDDLPLPBNARNYVRRVGLGVGRISTPFSVVGPREQT 420

Query: 421 NILESWWSNI 430  
 NILESWW+I  
 Sbjct: 421 NILESWWASI 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1486

A DNA sequence (GBSx1572) was identified in *S. agalactiae* <SEQ ID 4573> which encodes the amino acid sequence <SEQ ID 4574>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood	Transmembrane	30 - 46 ( 22 - 55)
INTEGRAL	Likelihood = -2.97	Transmembrane	110 - 126 ( 109 - 126)
INTEGRAL	Likelihood = -0.11	Transmembrane	89 - 105 ( 89 - 106)

----- Final Results -----

bacterial membrane	---	Certainty=0.4715 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 8823> which encodes amino acid sequence <SEQ ID 8824> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10  
 SRCFLG: 0  
 McG: Length of UR: 5  
 Peak Value of UR: 3.05  
 Net Charge of CR: 0  
 McG: Discrim Score: 4.64  
 GVE: Signal Score (-7.5): -1.66

Possible site: 36  
 >>> Seems to have a cleavable N-term signal seq.

Amino Acid Composition: calculated from 37  
 ALOM program count: 2 value: -2.97 threshold: 0.0

INTEGRAL	Likelihood	Transmembrane	100 - 116 ( 99 - 116)
PERIPHERAL	Likelihood = 1.38	56	

modified ALOM score: 1.09  
 icm1 HYPID: 7 CFP: 0.219

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane	---	Certainty=0.2190 (Affirmative)	< succ>
bacterial outside	--- <td>Certainty=0.0000 (Not Clear)</td> <td>&lt; succ&gt;</td>	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	--- <td>Certainty=0.0000 (Not Clear)</td> <td>&lt; succ&gt;</td>	Certainty=0.0000 (Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database and no corresponding DNA sequence was identified in *S. pyogenes*.

-1644-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1487

A DNA sequence (GBSx1573) was identified in *S.agalactiae* <SEQ ID 4575> which encodes the amino acid sequence <SEQ ID 4576>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0967 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1488

A DNA sequence (GBSx1574) was identified in *S.agalactiae* <SEQ ID 4577> which encodes the amino acid sequence <SEQ ID 4578>. This protein is predicted to be SgaT protein (sgaT). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.80 Transmembrane 441 - 457 ( 436 - 464)
INTEGRAL Likelihood = -7.64 Transmembrane 344 - 360 ( 339 - 376)
INTEGRAL Likelihood = -6.58 Transmembrane 403 - 419 ( 392 - 422)
INTEGRAL Likelihood = -6.48 Transmembrane 237 - 253 ( 235 - 261)
INTEGRAL Likelihood = -5.79 Transmembrane 105 - 121 ( 99 - 127)
INTEGRAL Likelihood = -5.52 Transmembrane 138 - 154 ( 137 - 155)
INTEGRAL Likelihood = -4.78 Transmembrane 18 - 34 ( 14 - 38)
INTEGRAL Likelihood = -2.97 Transmembrane 365 - 381 ( 365 - 383)
INTEGRAL Likelihood = -0.69 Transmembrane 41 - 57 ( 41 - 57)
INTEGRAL Likelihood = -0.16 Transmembrane 160 - 176 ( 160 - 176)

----- Final Results -----
bacterial membrane --- Certainty=0.4121 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC77150 GB:AE000491 orf, hypothetical protein [Escherichia coli K12]
Identities = 181/451 (40%), Positives = 274/451 (60%), Gaps = 25/451 (5%)

Query: 11 FSCNQLQNPAPFFVGLLVIGLYLLKKPLHDVFFAGFIKATVGYLLINVGAGLWTFRPTL 70
      ++ N ++G++ +GZ+LL+K + + G IK +G+++L G+G L +TF+P++
Sbjct: 30 FFCNVTNAPILLGIVTCLGYILLRKVSVLIRKTKITLIGFMLLQAGSGILSTFKFPVV 89

Query: 71 VALAKKENLEAAVIDPYPGLASANAKLBTNG-FISVATTALLIGPGINILLVALRKVTKV 129
      +++ + + A+ D Y AS A ++ M3 S A+L+ +N+ V LR++T +
Sbjct: 90 ANMSEVVGNGAISLITY---ASMMAITIDKNGAYSNGHVAIALLAINICVILLRITGI 146

Query: 130 RTLPFITCHIMVQCAATISVFVLLLIQLNLRFGAMAV----GLICLNVAWSSNNTVEAT 185
      RT+ +TGIM QQA I+V + + G+ N I+ LNW ++SM + T
Sbjct: 147 RTMLTGHIMFQAGLIAVTLFIF-----GYSMNTTICAILVSLYWGITSNNIKPT 200
```

-1645-

Query: 186 QRLTGGGGAIGHQQQFAIWVDKVAFFPGKKKEENLDNLKPTFLNIPHDTVVASATLML 245  
 Q +T G GF+IGHQQQFA W KVAPF GKKEE+++LKLIP +LNI PHD +V+ +A +M  
 5 Sbjct: 201 QEVTDGCGFSIGHQQQFASWIAYKVAFLGKKEESVEDLKLPGMLNIPHDNIVSTALVMT 260

Query: 246 VFFGGILAVLPGDIMSINKLIGFGAPVPTKQAFMYLLQTSITFSVYLFIIMQGMFVPT 305  
 +FPG IL GD + + K + +YLIQT +P+V +FI+ QGVRFV  
 Sbjct: 261 IFPGAILLSFGIDTVQ-----ANAGKHVHTVYILQTGFSPFAVAIFITIQGVRFVFA 311

Query: 306 ELTNAPQGISNKLPGSPFAVDVAASYGPGSSNAVLGSGAPGLIGQLITALLVFPKPI 365  
 EL+ AF GIS +L+PG+ A+D AA Y F + NAV+ GF +G IQQLI + +LV + I  
 10 Sbjct: 312 ELSEAPNGISQRLIPGAVALDCAITYSF-APNAVVMGPMWGTIGQLIAVGLIACGSSI 370

Query: 366 LIITGFVVFVDNAIAVYADKRGGKAAVAALSFISGITQVALGAVANGLLHAGGYHGN 425  
 LII GF+P+FF NA I V+A+ GGV+AA+ + + G+I++ AV L G++ + I  
 15 Sbjct: 371 LIITGFTPMFFSNATIGVFANHGGWRAALKICLVGMTEIFGCVNAVKLTMGS-ANWGM 429

Query: 426 IDPEFPMALRGYIFKILGIAGYVIVLCFFLA 456  
 D+ F +GIA ++ + LA  
 20 Sbjct: 430 ADNSILAPPMQGFSGIFPMVITIALA 460

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4579> which encodes the amino acid sequence <SEQ ID 4580>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -10.51 Transmembrane 441 - 457 ( 435 - 465)  
 INTEGRAL Likelihood = -7.80 Transmembrane 344 - 360 ( 339 - 376)  
 INTEGRAL Likelihood = -7.64 Transmembrane 238 - 254 ( 235 - 261)  
 30 INTEGRAL Likelihood = -5.63 Transmembrane 105 - 121 ( 100 - 127)  
 INTEGRAL Likelihood = -5.52 Transmembrane 138 - 154 ( 137 - 155)  
 INTEGRAL Likelihood = -5.20 Transmembrane 400 - 416 ( 392 - 422)  
 INTEGRAL Likelihood = -4.78 Transmembrane 18 - 34 ( 14 - 39)  
 INTEGRAL Likelihood = -2.97 Transmembrane 365 - 381 ( 365 - 383)  
 INTEGRAL Likelihood = -1.49 Transmembrane 160 - 176 ( 160 - 177)  
 35 INTEGRAL Likelihood = -0.53 Transmembrane 41 - 57 ( 41 - 57)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5203 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 40 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>P:AC77150 GB:AE000491 orf, hypothetical protein [Escherichia coli]  
 45 Identities = 182/461 (39%), Positives = 279/461 (60%), Gaps = 25/461 (5%)

Query: 1 MEMLAPLIMFSONILQNPAPFVGLVILIGYLLKKPIYVFGGVKATVGYLHIVCGG 60  
 ME+L F ++ N +G++ +GV+LL+K + + G +K +G++L G+G  
 Sbjct: 20 MEILNITPTVFPNQVMINAPLLGIVTCLGYILLKESVSVIKGTIKTIIGPMLQASS 79

50 Query: 61 GLVITFPFVLVAKKFELGAVIDPYGLAANTKLEMG-PISVATALLIGFVNI 119  
 L +TF+P++ +++ + + A+ D Y + A ++ MG S A+L+ +NI  
 Sbjct: 80 ILTSTFPVVKMSVYGINGAISDTYASMMAT---IDRMGRAYSWMGYAVLLALALNIC 136

55 Query: 120 LVALKRKVKTKVLTIFTHIMVQQAATISVPLLIPQPNAGAWAV---GHCGLYWA 175  
 V LR++T +RT+ +THIM QQA I+V + + + W I+ L+VW  
 Sbjct: 137 VLLLRITGIRTIMLGHIMFQAGLIAVTLFIP-----GYEMTTITICAILVSLVWG 190

60 Query: 176 ISSNTVEATQRLTGGGGAIGHQQQFAIWVDKVAFFPGKKKEENLDNLKPTFLNIPHD 235  
 I+SNM + +Q +T G GF+IGHQQQFA W KVAPF GKKEE+++LKLIP +LNI PHD  
 Sbjct: 191 ITSNWYKPTQEVTDGCGFSIGHQQQFASWIAYKVAFLGKKEESVEDLKLPGMLNIPHD 250

Query: 236 TVVASATLMLVFGGAILAVLPGDIMSVDLIGFGAPNPAKQAFMYLLQTSITFSVYLFI 295  
 +V+ +A +M +FPGAIL GD + + K + +YLIQT +P+V +FI  
 65 Sbjct: 251 NIVSTALVMTIFPGAILLSFGIDTVQAM-----ACKKHVHTVYILQTGFSPFAVAIFIT 301

-1646-

Query: 296 INQGVRFVSELTNAFQGISSKLLPGSFPAVDVAASYGPGSSNAVLGSGFAGLIGQLITI 355  
 + QGVRFV+EL+ AF GIS +L+PG+ A+D AA Y F + NAV+ GF +G IGQLI +  
 5 Sbjct: 302 ITQGVRFVVAELSEAFNGISQRLIPGVALIDCAALYSF-APNAVVGSGFWGTIGQLIAV 360

Query: 356 ALLVIFIQNFILITGFVPVFFDAAIAVADKRGKGAVALSFISGIIQVALGAVAVGL 415  
 +IV + ILII GF+P+FF NA I V+A+ GSW+AA+ + + G++++ AV L  
 Sbjct: 361 GILVAGGSSILIIIGPIFMFSGNATIGVFANHGPGWRAALICLVMNIEIGCVAVAVKL 420

10 Query: 416 LGLTGGYHGNIDLVLFWLPPGYLKFPLGSIAGYVLCIFLLA 456  
 G++ + G D + P F +GIA ++ + LA  
 Sbjct: 421 TGMS-ANWGMADWSILAPFMGPFSGISIAPMAVIIVIALA 460

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 437/476 (91%), Positives = 457/476 (95%)

Query: 1 MENFLAPLNNFSONILQNPFAFFVGLVLVLIGYLLKKPLHDVFAGFIKATVGYLILNVGAG 60  
 ME LAPLNNFSONILQNPFAFFVGLVLVLIGYLLKKP+++VFAGF+KATVGYLILNVGAG  
 20 Sbjct: 1 MENLLAPLNNFSONILQNPFAFFVGLVLVLIGYLLKKPIYEVFAGFVKATVGYLILNVGAG 60

Query: 61 GLVNTFRFILVALAKKFNLEAAVIDPYFGLASNAKLETMCFISVATTALLIGFGINILL 120  
 GLV TFRFILVALAKCF L-AAVIDPYFGLA+AN KLE MGPISVATTALLIGFG+NIL  
 Sbjct: 61 GLVNTFRFILVALAKKFELKAAVIDPYFGLAANAKLEMGFISVATTALLIGFGVNIIL 120

25 Query: 121 VALRKVKVTRLFTITGHIMVQQAATISVFLVLLIPQLRNGFGAMAVGIIIGLYWAVSSNM 180  
 VALRKVKVTRLFTITGHIMVQQAATISVFLVLLIPQ +N FGAMAVGIIIGLYWA+SSNM  
 Sbjct: 121 VALRKVKVTRLFTITGHIMVQQAATISVFLVLLIPQFNAPGAMAVGIIIGLYWAISNM 180

30 Query: 161 TVEATQRLTGGGGFAIGHQQQFAINFVDKVAFFPGKKEBNLDNKLFTFNI FHDTVAS 240  
 TVEATQRLTGGGGFAIGHQQQFAINFVDKVAFFPGKKEBNLDNKLFTFNI FHDTVAS  
 Sbjct: 161 TVEATQRLTGGGGFAIGHQQQFAINFVDKVAFFPGKKEBNLDNKLFTFNI FHDTVAS 240

Query: 241 ATMLNVFFGGILAVLGPDIMSIVLIGPGAFVFTKQAFPMYILQTSLTSFVYLFIAMGV 300  
 ATMLNVFFG ILAVLGPDIMS+V LIGPGAF F EQAFPMYILQTSLTSFVYLFIAMGV  
 35 Sbjct: 241 ATMLNVFFGAILAVLGPDIMSVDLIGPGAFNPAKQAFPMYILQTSLTSFVYLFIAMGV 300

Query: 301 RMFVTELTNAFQGISKNLLPGSFPAVDVAASYGPGSSNAVLGSGFAGLIGQLITIALIV 360  
 RMFV+ELTNAFQGIS+KLLPGSFPAVDVAASYGPGSSNAVLGSGFAGLIGQLITIALIV+  
 40 Sbjct: 301 RMFVSELTNAFQGISKNLLPGSFPAVDVAASYGPGSSNAVLGSGFAGLIGQLITIALVI 360

Query: 361 FKMPILIIITGFVPVFFDAAIAVADKRGKGAVALSFISGIIQVALGAVAVGLIGL 420  
 FKMPILIIITGFVPVFFDAAIAVADKRGKGAVALSFISGII+QVALGAVAVGLIGL G  
 Sbjct: 361 FKMPILIIITGFVPVFFDAAIAVADKRGKGAVALSFISGIIQVALGAVAVGLIGL 420

45 Query: 421 GYHGNIDFEFFMLAPGYIFKYLGLAGVYVLCFFLAIPQLQFMKSKDKRAYYRGA 476  
 GYHGNID FWL FG+FK+LGLAGV+VC+P LAIPQLQF K+KDKRAYYRGA  
 Sbjct: 421 GYHGNIDLVLFWLPPGYLKFPLGSIAGYVLCIFLLAIPQLQFQAKAKDKRAYYRGA 476

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1489

A DNA sequence (GBSx1575) was identified in *S. agalactiae* <SEQ ID 4581> which encodes the amino acid sequence <SEQ ID 4582>. Analysis of this protein sequence reveals the following:

55 Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.1225 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>



-1647-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG34743 GB:AE000033 similar to PTS system: EIIB (Mycoplasma pneumoniae)  
Identities = 40/89 (44%), Positives = 62/89 (68%), Gaps = 1/89 (1%)

5 Query: 4 VLTACGNGGSSMVIRKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNHLE 63  
++ ACGNGG+SM+IK+KVE +++IG + A S+G+ R3+ + DI+++S HL E  
10 Sbjct: 8 ITAACGNGGTSMLIKIKVKEIKKELGYTRKVEA-LSNGQTGNGHSADIISSIHLSR 66  
Query: 64 LDGRTKGHLVGLDNLMDNEIKTKLQEL 92  
+ K +VG+ NEMD+NEIK L +tL  
20 Sbjct: 67 FNPRAKAKIVGVNLMDNEIKTKLQEL 95

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4583> which encodes the amino acid sequence <SEQ ID 4584>. Analysis of this protein sequence reveals the following:

15 Possible site: 42  
>>> Seems to have no N-terminal signal sequence  
----- Final Results -----  
20 bacterial cytoplasm --- Certainty=0.0977 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 85/92 (92%), Positives = 90/92 (97%)  
25 Query: 1 MVKVLITACGNGGSSMVIRKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNH 60  
MVKVLITACGNGGSSMVIRKVENALRQLGV++ +SASCSVGEAKGLA+ YDIVVASNH  
Sbjct: 1 MVKVLITACGNGGSSMVIRKVENALRQLGVTDQSASCSVGEAKGLASGYDIVVASNH 60  
30 Query: 61 IHELDGRTKGHLVGLDNLMDNEIKTKLQEL 92  
IHELDGRTKGHLVGLDNLMDNEIKTKLQEL  
Sbjct: 61 IHELDGRTKGHLVGLDNLMDNEIKTKLQEL 92

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1490

A DNA sequence (GBSx1576) was identified in *S.agalactiae* <SEQ ID 4585> which encodes the amino acid sequence <SEQ ID 4586>. This protein is predicted to be a pentitol phosphotransferase enzyme ii, a component (ptxA). Analysis of this protein sequence reveals the following:

40 Possible site: 38  
>>> Seems to have no N-terminal signal sequence  
----- Final Results -----  
45 bacterial cytoplasm --- Certainty=0.3309 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77152 GB:AE000491 putative PTS system enzyme II A component  
[Escherichia coli K12]  
Identities = 64/150 (42%), Positives = 97/150 (64%), Gaps = 2/150 (1%)  
50 Query: 1 MNLKAPFIENDSIRLKLGSADWKEAIKLSIDPLIESGAVDAAEYDAITESTREFGPPYIL 60  
M L+ + EN SIRL+ A W+EA+K+ +D L+ + V+ YY AI++ E+RGPEY++  
55 Sbjct: 1 MKLRDSLAENKSIRLQAEFTWQRAVKIGVDLLVAADVPEFYQAILDGVQCGPPYVI 60  
Query: 61 MFGMMPHARPENGVKRDAFSLITLITPEVVF--PDGKEVSVLLAAATSSAHTSVAIPQ 118

-1648-

PG+AMFH RPE GVK+ FSL+TL +P+ F D V +L+ +AA + H V I Q  
 Sbjct: 61 APGLAMFHGRPEEGVKCTGFSLVTLKKPLEFNEENDPVDILITMAAVDANTHGEVGINQ 120

Query: 119 IIALFELENSIQRLTECCQAEKVLAMVEES 148  
 I+ LEE E + RL C+ +EVL +++ +  
 Sbjct: 121 TVNLFEEDEBNFDRLRACRTGCEVLDLIDRT 150

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4587> which encodes the amino acid sequence <SEQ ID 4588>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2287 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 113/161 (70%), Positives = 137/161 (84%)

Query: 1 MNLKQAFITENDSIRLKLGSADWKAIKLSIDPLIESGAVDAEYYDAIESTEEFGPPYIL 60  
 MNLKQAFI+N+SIRL LSA W+R++L++ FLI+S AV + YYDAII STE++GFYY+L  
 Sbjct: 1 MNLKQAFITDNNIRLKLGSADTQCAVRLAVQFLIDSKAVTSAYDAI IASTEKYGFYYVL 60

Query: 61 MPGMMPHARPEAGVVRDAFSLITLTFVFPDGKEVSVLLALAAATSSAHTSVAIFQII 120  
 MPGMMPHRA GV R+AF+LITLT+V F DGKEVSVLL LAAT +IHT+VAIFQI+  
 Sbjct: 61 MPGMMPHRAEAGLGVRNRAFALITLTKFVTSDGKEVSVLLITLAATDPSIHTTVAIFQIV 120

Query: 121 ALFELENSIQRLTECCQAEKVLAMVEESKNSPYLEGDLDES 161  
 ALFEL+N+I+RL CQ KEVL MVEESK+SPYLEG+DL +  
 Sbjct: 121 ALFELNATIERLVACQSPKEVLAMVEESKDSPLYEGMDLNA 161

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1491

A DNA sequence (GBSx1577) was identified in *S.agalactiae* <SEQ ID 4589> which encodes the amino acid sequence <SEQ ID 4590>. This protein is predicted to be probable hexulose-6-phosphate synthase. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1504 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77153 GB:AF000491 probable hexulose-6-phosphate synthase  
 [Escherichia coli K12]  
 Identities = 108/217 (49%), Positives = 141/217 (64%), Gaps = 3/217 (1%)

Query: 5 LPNLQVALDHSIDLQGAIAAASVGHSEVDTIAGTVCILQVGSRLVFLRSLFPFKIIVAD 64  
 LP LQVALD+ + A + + EVD+IE GT+ + G V L+L+P KI+AD  
 Sbjct: 3 LPLQLVALDNTQWDSAYETTRILIAEVDITIEVGTILCVGSEVRAVRDLKALYPHKIIVAD 62

Query: 65 TKCADAGCTVAKNNVARGADWTCITCCATITMEALAKAIKEERDGRGEIQIELYGDWY 124  
 K ADAG +++ ADW+T ICA I T + AL KE GD +QIEL G WT+  
 Sbjct: 63 AKIADAGKILSRMCFANADWYTVICADINIRGALDVAKEPKND--VQIKELTYGWTW 119

-1649-

Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKLLIDMGFRVSVTGGGLSTDTLQL 184  
 EQAQQW DGI Q +YH+SRDA AG WGE D+ +K+L DMGF+V+VTGGL+ + L L  
 Sbjet: 120 EQAQQWLDAGIGQVYTHSRDAQAAGVAMGEADITAIKRLSDMGFPKVTVTGGLAEDLPL 179

Query: 185 FBGVVDVFTFIAGRGITEADDPAAARAFKDEIKRIMG 221  
 F+G+ + FIAGR I +A P AAR FK I +WG  
 Sbjet: 180 FBGPIHVFIAGRBIRDAASPVEARQFKRSIAELWG 216

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4591> which encodes the amino acid sequence <SEQ ID 4592>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.1473 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 206/217 (94%), Positives = 212/217 (96%)

- Query: 5 LPNLQVALDHSDLQGAIKAAVSVGHEVDVIEAGTVCLLQVGSSELVEVLRSLFPFKIIVAD 64  
 +PNLQVALDHSDLQGA+KAAV+VGHEVDVIEAGTVCLLQVGSSELVEVLRSLFP+KIIIVAD  
 Sbjet: 4 LPNLQVALDHSDLQGAIVAAVVGHEVDVIEAGTVCLLQVGSSELVEVLRSLFPFKIIVAD 63
- Query: 65 TKCADAGGTAVAKNNAVAGADWMTCCICCATIPTMEALKAKEERDGRGEIQIELYGDWITY 124  
 TKCADAGGTAVAKNNA RGADWMTCCICCATIPTMEALKAKEERDGRGEIQIELYGDWITY  
 Sbjet: 64 TKCADAGGTAVAKNNAKRGADWMTCCICCATIPTMEALKAKEERDGRGEIQIELYGDWITY 123
- Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKLLIDMGFRVSVTGGGLSTDTLQL 184  
 EQAQ WLDAGISQAIYHQSRDALLAGETWGEKDLNKVK LIDMGFRVSVTGGGL DTL+L  
 Sbjet: 124 EQAQLNLDAGISQAIYHQSRDALLAGETWGEKDLNKVKTLIDMGFRVSVTGGGLDVTLLRL 183
- Query: 185 FBGVVDVFTFIAGRGITEADDPAAARAFKDEIKRIMG 221  
 FBGVVDVFTFIAGRGITEA+DPAAARAFKDEIKRIMG  
 Sbjet: 184 FBGVVDVFTFIAGRGITEADDPAAARAFKDEIKRIMG 220

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1492

A DNA sequence (GBSx1578) was identified in *S.agalactiae* <SEQ ID 4593> which encodes the amino acid sequence <SEQ ID 4594>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.4179 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22686 GB:U32783 hexulose-6-phosphate isomerase, putative  
 [Haemophilus influenzae Rd]

- Identities = 143/282 (50%), Positives = 199/282 (69%), Gaps = 3/282 (1%)

Query: 5 IGIYEKATPGHFWLERLQPAKELGPDFVFLSIDESDERLARLSWSEKRELVLKAIPTET 64  
 IGIYEKA PK+ W ERLL AK GP+F+8+SIDBS+RL+EL W+K ER+ L ++I ++  
 Sbjet: 6 IGIYEKALPKNITWQRLSLAKAGQFRTFMSIDBSNRLSLNLTWKTERTALHQSGITQS 65

-1650-

Query: 65 GVRVPTITTS~~SGHRRFTWGSNNPEKEARAMDMMKKCI~~VFAQDIGIRNIQLAGYDYYEKS 124  
 G+ +P++ S HRRFF GS + + ++ +M+K I + ++GIR IQLAGDYDYTE++  
 Sbjct: 66 GTITPSMCL~~SHRRFFFGSGDKKIRKQSF~~IMEKAIDLSVNLGIRITQLAGDYDYTEKQD 125

Query: 125 PETRARFINKLRQACTWAEAAQVILSIEIMDDPFMSIEKYLAVEKEIDSPFLVFDY 184  
 ET F + + A T A AOV L+ +EIMD FFM+SI ++ + I+SP+ VFYD G  
 Sbjct: 126 EETIKYFPGIRKFAVTLAASQVTLAVEIMUTPFMSISRWKKMUTLINSFWPTFVYDYG 185

Query: 185 NVSAWNLMS~~EFYNGHRSIALHLKDT~~YAVTETSGQFRDVPFGQGVDEWHFAVIKK 244  
 N-SAWN++ E G I+A+H+KDTY VIETSGQFRDVPFGQGVCD+ F+++KK  
 Sbjct: 186 NL~~SAWN~~NIEEELTLGDKISALHLKDTYVTETSGQFRDVPFGQGVDFVHF~~FL~~LKK 245

Query: 245 TTYNGPFLIEWNS~~ENCETVEETRAAIKEAQD~~FLYLMKXTGV 286  
 NY G FLIEW+E EE I +A+ ++ MEK G+  
 Sbjct: 246 IAYRGAFLIEWNTEK---NEEFLLEIIQARNTVQ~~Q~~MEKAGL 284

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4595> which encodes the amino acid sequence <SEQ ID 4596>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1489 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 240/286 (83%), Positives = 271/286 (93%)

Query: 1 MTRPIGIYKATPK~~FNMLERLQFAKELGDFV~~ELSIDSDERLARLEWSKKEERL~~LVKA~~ 60  
 M RPIGIYKATPK F W ERLQFAK+LGDFVE+S+DES RLARLEW+KEERL+LVKA  
 Sbjct: 15 MARPIGIYKATPKQFTMRERLQFAKDLGDFVEMSVDES~~DARLARLEWTKERL~~DLVKA 74

Query: 61 IFETGVRVPTITTS~~SGHRRFTWGSNNPEKEARAMDMMKKCI~~VFAQDIGIRNIQLAGDYDYY 120  
 I+ETG+R+PTI FSGHRR+F+GSN+F EA+++ +MK+CI AQD+G+R IQLAGDYDYY  
 Sbjct: 75 IYETGIRITPTI~~CFSGHRRYPLSGNDPAIEAKSLKLMKQC~~IELAQDLGVRTIQLAGDYDYY 134

Query: 121 EKS~~PETRARFINKLRQACTWAEAAQVILSIEIMDDPFMSIEKYLAVEKEIDSPFLVY~~ 180  
 E+AS~~PETRARFINKLRQ~~+C WAEAAQV+LSIEIMDDF+NSIEKYLAVEKEIDSPFLVY  
 Sbjct: 135 EK~~SPETRARFINKLRQ~~SCDWAEAAQVILSIEIMDDPFMSIEKYLAVEKEIDSPFLVY 194

Query: 181 PD~~TGNVSAWNLMS~~EFYNGHRSIALHLKDTYAVTETSGQFRDVPFGQGVDEWHFA 240  
 PD GNVSAWNLMS~~EFYNGH~~+SIALH+KDTYAVTETSGQFRDVPFGQGVCDW+E+FA  
 Sbjct: 195 PD~~AGNVSAWNLMS~~EFYNGHRSIALHLKDTYAVTETSGQFRDVPFGQGVCDW~~ELFA~~ 254

Query: 241 VIKKINYNGPFLIEWNS~~ENCETVEETRAAIKEAQD~~FLYLMKXTGV 286  
 V+KKINYNGPFLIEWNS~~ENC~~+VTEET+AAIK~~EAQD~~FLYLMK G+  
 Sbjct: 255 VLKKNYNGPFLIEWNS~~ENC~~IVETKA~~AIKEAQD~~FLYLMK~~AGL~~ 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1493

A DNA sequence (GBSx1579) was identified in *S.agalactiae* <SEQ ID 4597> which encodes the amino acid sequence <SEQ ID 4598>. This protein is predicted to be L-ribulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

-1651-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2559 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase  
 [Bacillus stearothermophilus]  
 Identities = 143/229 (62%), Positives = 176/229 (76%), Gaps = 2/229 (0%)

Query: 5 LQEMRERVCERANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVDL 64  
 L+E+++ V EAN LP + LV FTWGNVS +DRE GL+VIKPSGV YD+LT ++MV DL  
 Sbjct: 2 LSELQAVLEARNLQLPQYRIVITFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDL 61

Query: 65 EGNIVEGDLNPSSDLPPTHVQLYKAWPEVGGIVHTHSTEAVGWAQAGRDIPFYGTTHADYF 124  
 (N+VEGD) PSSD PTH+ LYK +P +GGIVHTST A WAQAG+ IP GTTHADYF  
 Sbjct: 62 TGNVVEGDLNPSSDLPPTHMLYKQFPFGIGIVHTHSTWATVWAQAGKGI PALGTTHADYF 121

Query: 125 YGFVPCARSLSEDEVNTAYEKTGSVIIIEPERRDLDPMAVGVIVRNHGPTTWGKDPQA 184  
 YG +PC R ++ +E+ AYE ETG VI E F R LDP+ +EG+V HGPF WSKDPA  
 Sbjct: 122 YGEIFCTPRFTNBSIQGAYLESTQKVTETET--RFLDPLQMPGLVHGHPFAMGKDPAN 179

Query: 185 AVYHSVVLLEEVAKMNRFTQINPRVEPAPKYIMDKHYLRKHGPNAYYQK 233  
 AV+++VVLLEEVAKM T +NP +P + +D+HYLRKHGPNAYYQK  
 Sbjct: 180 AVENNVVLLEEVAKMAARTYMLNPAKPISTQLDRHYLRKHGPNAYYQK 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4599> which encodes the amino acid sequence <SEQ ID 4600>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2257 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 207/234 (88%), Positives = 220/234 (93%)

Query: 1 MAKSLQENKERVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPESIMV 60  
 MAK+LQENKERVCEANKSLP H LKFTWGNVSEV RE G IVIKPSGVDYD LTPEIMV  
 Sbjct: 1 MAKNLQENKERVCEANKSLPQHLVKFTWGNVSEVCEGLGRIVIKPSGVDYDLLTPEIMV 60

Query: 61 VTDLGNIVEGDLNPSSDLPPTHVQLYKAWPEVGGIVHTHSTEAVGWAQAGRDIPFYGTTH 120  
 VTDL+GN+VSGDLNPSSDLPPTHV+LYKAWPEVGGIVHTHSTEAVGWAQAGRDIPFYGTTH  
 Sbjct: 61 VTDLGNVVEGDLNPSSDLPPTHVLYKAWPVGIVHTHSTEAVGWAQAGRDIPFYGTTH 120

Query: 121 ADYFYGVPCARSLSEDEVNTAYEKTGSVIIIEPERRDLDPMAVGVIVRNHGPTTWGK 180  
 ADYFYGVPCARSL++ EV+ AYE+ETG+VI+KEF +R LDPMAVGVIVRNHGPTTWGK  
 Sbjct: 121 ADYFYGVPCARSLTAEVDGAYEQETGNVILEEFSKRGLODPMAVGVIVRNHGPTTWGK 180

Query: 181 DPAQAVYHSVVLLEEVAKMNRFTQINPRVEPAPKYIMDKHYLRKHGPNAYYQK 234  
 P QAVYHSVVLLEEVA+MNR TEQINPRVEPAP+YIMDKHYLRKHGPNAYYQK  
 Sbjct: 181 TPBQAVYHSVVLLEEVAKMNRFTQINPRVEPAPRYIMDKHYLRKHGPNAYYQK 234

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1652-

**Example 1494**

A DNA sequence (GBSx1580) was identified in *S. agalactiae* <SEQ ID 4601> which encodes the amino acid sequence <SEQ ID 4602>. This protein is predicted to be transaldolase (tal). Analysis of this protein sequence reveals the following:

5       Possible site: 45  
       >>> Seems to have no N-terminal signal sequence

10       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.4232 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10149> which encodes amino acid sequence <SEQ ID 10150> was also identified.

15       The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB98962 GB:U67539 transaldolase [Methanococcus jannaschii]  
 Identities = 124/214 (57%), Positives = 157/214 (72%)

20       Query: 19 MKYFLDTADVSEIRRLNRLGIVDGVTTNPTLISRBGRDPKEVINEICQIVDGPVSAEVTG 78  
           NK+FLDTA+V EI++ LG+VDGV+TNPT+++BGRDF EV+ EIC+IV+GPVSAEV  
       Sbjct: 1 MKFFLDTANVESIKKTAELGLVDGVTTNPTLVAKEGRDFYEVVKEICRIVEGSPVSAEVTG 60

      Query: 79 LTCDEMVTREARIKNSPNVVKIPMTSEGLAANVQLSKSEGIKTNVILIFTVAQGLSAMK 138  
           + MV EARE+AK + N+V+KIPMT++G+ AV LS EGIKINVTIL+F+ Q L A K  
       Sbjct: 61 TDAGNVKEARELAKLADNIVIKIPMTYOGMKAVKILSAEGIKINVTILVFSPQLQALVAK 120

25       Query: 139 AGATYISPFVGRLEDIGTDAYALIRDLRHIIDFYGQSEIIAASIRGLAHVGVAKCGAH 198  
           AGAT+SPFVGRLE+DIG LI D+ I Y ++E+T AS+R HV AK GA  
       Sbjct: 121 AGATYISPFVGRLEDIGTHVGMKLEDVVKIYKNDIKTEVIVASVRHPHWLEAAKIGAD 180

30       Query: 199 IATIPDKTFASLFTPLFDKGIETFLKDWDSFKK 232  
           IAT+P LF HPLED G+E FLKDW+ K  
       Sbjct: 181 IATMPPAVMDKLFNHPLEIDGLERFLKDWDEYLN 214

35       A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4603> which encodes the amino acid sequence <SEQ ID 4604>. Analysis of this protein sequence reveals the following:

      Possible site: 15  
       >>> Seems to have no N-terminal signal sequence

40       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1902 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45       An alignment of the GAS and GBS proteins is shown below.

Identities = 162/214 (75%), Positives = 180/214 (83%)

      Query: 19 MKYFLDTADVSEIRRLNRLGIVDGVTTNPTLISRBGRDPKEVINEICQIVDGPVSAEVTG 78  
           NK+FLDTA+V I+ +N LG+VDGVTTNPTLISRBGRDP+ VI EIC IVDGP+SAEVTG  
       Sbjct: 1 MKFFLDTANVAIKANLGLVVDGVTTNPTLISRBGRDPFETVINEICQIVDGPVSAEVTG 60

      Query: 79 LTCDEMVTREARIKNSPNVVKIPMTSEGLAANVQLSKSEGIKTNVILIFTVAQGLSAMK 138  
           LT D MV EAR IAKN NVVVKIPMT EGI A + LSKSEGIKTNVILIFTV+QGL AMK  
       Sbjct: 61 LTADAMVEARSIKAWHDNIVVKIPMTSEGLAKATNLSKSEGIKTNVILIFTVSGQLAMAMK 120

55       Query: 139 AGATYISPFVGRLEDIGTDAYALIRDLRHIIDFYGQSEIIAASIRGLAHVGVAKCGAH 198  
           AGAT+ISPF+GRLEDIGTDAY LI DLR IID Y PQ+RITTAASIR AKVE VAK GAH  
       Sbjct: 121 AGATYISPFVGRLEDIGTDAYQLISDLREITIDVDFQABIIAASIRTAHVEAVAKIGAH 180

-1653-

Query: 199 IATIPDKTFASLFTPLTNGIETFLKDWDSFKK 232  
 IATIPD FA + HFLT G++TF++DW SFFK  
 Sbjct: 181 IATIPDPLEAKMTQHPLATNGLKTFMEDWASFKK 214

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1495

A DNA sequence (GBSx1581) was identified in *S.agalactiae* <SEQ ID 4605> which encodes the amino acid sequence <SEQ ID 4606>. Analysis of this protein sequence reveals the following:

10 Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1263 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAB14129 GB:Z99115 transcriptional regulator (LacI family)  
 [Bacillus subtilis]  
 Identities = 108/331 (32%), Positives = 188/331 (56%), Gaps = 12/331 (3%)

Query: 6 TISDIANLVGSKIATVSYLNGRYKIMSLQTKKIRLAIKETGYQPSKIAQSLVTCVIRT 65  
 TI D+A GVSK+TVS Y+NG +S + + I+ AI E Y+PSK+AQ L K ++  
 25 Sbjct: 10 TIKDVABCAQVSKSTVSYKINGKIDAISSPEKVINIKKALAEINLYRPSKMAQGLKIKKSKL 69

Query: 66 IGVVIADITNPFISVMKGIHDTQQQGYGVNFTNSDNDIDISLENIARLNQNVSGIIL 125  
 IG V+ADITNPF + +G+ C Q+GYS+ N+DN + E E L +L +V G+IL  
 30 Sbjct: 70 IGVVVIADITNPFISVAAFRGVEVCDQYGYSIMVCTNDSPEKREMLLKLEAHSVSGIIL 129

Query: 126 DSVDPNHSFIETLSNDRL--VMVDRAQDKIKVDITVASDNKESTQIFLEKQAEAGYHDIYF 183  
 ++ N + + ++ ++DR+ D+K+DTV +DN+ T+ L+K+ GY D+  
 35 Sbjct: 130 NATGENKDVLRFAEQQIPTILIDRLKLPDLKLDVTITDNRNITREILQKVYSKYTDVAL 189

Query: 184 VTYPIBEISTRELRYBGFKEVVS-SNPDKLIITE-DGSTQRILDI-----IEHSBQKP 235  
 T PI IS R R ++E+ S N + L + E D + L E EQK  
 40 Sbjct: 190 FTEPISSISPRAEAAVYQMAVGVNGLVRLRHEIDVDKDBQLKAEIRSPHKMEBQKK 249

Query: 236 GFLMMNSFTLLNPMKINQSTVSYPEDYGLGSYBDEWVQLTPNVSCIKQDSYGIGCLA 295  
 L +NG +L + + + + P+D G+ ++D EW ++P ++ I Q S+ +G A  
 45 Sbjct: 250 AILALNGLIMLKITSCESELGLRIPQDQIGIAGVDTEWYKLIQPGITTTIAQPSHMGRTA 309

Query: 296 AQCLIEKISQONEPTTARLLEVNQIVIRQS 326  
 + ++++++ + + +E+ +++++R+S  
 50 Sbjct: 310 MERVLRRIE--GDGAPQTTELEAKVIMRKS 338

There is also homology to SEQ ID 2366.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 50 Example 1496

A DNA sequence (GBSx1582) was identified in *S.agalactiae* <SEQ ID 4607> which encodes the amino acid sequence <SEQ ID 4608>. Analysis of this protein sequence reveals the following:

55 Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1654-

bacterial cytoplasm --- Certainty=0.1661 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1497

10 A DNA sequence (GBSx1583) was identified in *S.agalactiae* <SEQ ID 4609> which encodes the amino acid sequence <SEQ ID 4610>. This protein is predicted to be GLYCERATE DEHYDROGENASE. Analysis of this protein sequence reveals the following:

Possible site: 29  
 >>> Seems to have an uncleavable N-term signal seq

15

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB50351 GB:A240287 GLYCERATE DEHYDROGENASE [Pyrococcus abyssi]  
 Identities = 123/325 (37%), Positives = 192/325 (58%), Gaps = 8/325 (2%)

25

Query: 1 MDKKKILVGIIVPKBOLRLMDRFDVTSYSED-RPPSRDYVLEHLSEYDQWLLM-GQKGDK 58  
 M K ++ +T +P+ G+ L F+V ED R R+ +LE + + D + M ++ D+  
 Sbjct: 1 MSKPRVFITREIPEVGIEMLEKEFEVEWEDEIREIPREILLEKVKVDALVTMLSERIDR 60

30

Query: 59 EMIDAGENLIQIISNAVGFHDVDTAYAKEGIIIVSNPQAVRVPTAEMTFALILAAKRL 118  
 E+ + L+I++ AVG+D++D A ++GI V+N+P + T++ FAL+LA ++ L  
 Sbjct: 61 EVFERAPRLRIVANYAVGYNDIVVEEATKRGIIYVNTPGVLITATADLAPALLLATARHL 120

35

Query: 119 AFYDSIVRSGEW----IDPSEQRYQGLTLQGSLGIYMGRIGLTVANFAKAFGMTVVTN 174  
 D RSGEW + + + G + G T+GI G GRIG +A A+ F M ++Y  
 Sbjct: 121 VKSDKFRSRSGEWMKGRVAMHPKFWLGYDVVKGKTIIGIIPGRIOQAIAKRARGFDMRILY 180

40

Query: 175 DVYRLPEDEKEKELGVLTLEFDQLIKTADVITIHAPALSTIHKFNKDVFAKMKRNSYLIN 234  
 R PE EKEL + D+L++ +D + + P T H N++ MK + LIN  
 Sbjct: 181 SRTKRFV-VKELNASEFKLEDELLRSDDFVLAFLANKETVHMINERLKMGMRTAILIN 239

45

Query: 295 TLAEEADNIIAPFDGK-PQIVNK 318  
 +A+ A+N+IAF G+ P +VN+  
 Sbjct: 300 GMAKIVARNLIAFKRGEVPTLVNR 324

There is also homology to SEQ ID 124.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



-1655-

**Example 1498**

A DNA sequence (GBSx1585) was identified in *S.agalactiae* <SEQ ID 4611> which encodes the amino acid sequence <SEQ ID 4612>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1898(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
15 vaccines or diagnostics.

**Example 1499**

A DNA sequence (GBSx1586) was identified in *S.agalactiae* <SEQ ID 4613> which encodes the amino acid sequence <SEQ ID 4614>. This protein is predicted to be PTS system, galactitol specific IIC component. Analysis of this protein sequence reveals the following:

```

20  Possible site: 25
    >>> Seems to have no N-terminal signal sequence

    INTEGRAL Likelihood = -13.27 Transmembrane 254 - 270 ( 245 - 277)
    INTEGRAL Likelihood = -9.24 Transmembrane 77 - 93 ( 71 - 100)
    INTEGRAL Likelihood = -9.24 Transmembrane 367 - 383 ( 364 - 386)
25  INTEGRAL Likelihood = -8.28 Transmembrane 32 - 48 ( 26 - 54)
    INTEGRAL Likelihood = -7.38 Transmembrane 186 - 202 ( 182 - 215)
    INTEGRAL Likelihood = -6.26 Transmembrane 158 - 174 ( 151 - 180)
    INTEGRAL Likelihood = -5.79 Transmembrane 279 - 295 ( 276 - 296)
    INTEGRAL Likelihood = -1.12 Transmembrane 342 - 358 ( 342 - 359)
30  INTEGRAL Likelihood = -0.00 Transmembrane 308 - 324 ( 308 - 324)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8825> which encodes amino acid sequence <SEQ ID 8826> was also identified. Analysis of this protein sequence reveals the following:

```

40  Lipop: Possible site: -1 Crend: 9
    McG: Discrim Score: 8.30
    GVH: Signal Score (-7.5): 2.97
    Possible site: 58
    >>> Seems to have a cleavable N-term signal seq.
    ALOM program count: 9 value: -13.27 threshold: 0.0
45  INTEGRAL Likelihood = -13.27 Transmembrane 321 - 337 ( 312 - 344)
    INTEGRAL Likelihood = -9.24 Transmembrane 144 - 160 ( 138 - 167)
    INTEGRAL Likelihood = -9.24 Transmembrane 434 - 450 ( 431 - 453)
    INTEGRAL Likelihood = -8.28 Transmembrane 99 - 115 ( 93 - 121)
    INTEGRAL Likelihood = -7.38 Transmembrane 253 - 269 ( 249 - 282)
50  INTEGRAL Likelihood = -6.26 Transmembrane 225 - 241 ( 218 - 247)
    INTEGRAL Likelihood = -5.79 Transmembrane 346 - 362 ( 343 - 363)
    INTEGRAL Likelihood = -1.12 Transmembrane 409 - 425 ( 409 - 426)
    INTEGRAL Likelihood = -0.00 Transmembrane 375 - 391 ( 375 - 391)
    PERIPHERAL Likelihood = 0.69 188
55  modified ALOM score: 3.15

```

-1656-

\*\*\* Reasoning Step: 3

----- Final Results -----

- 5                   bacterial membrane --- Certainty=0.6307(Affirmative) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- >GP:BAB03909 GB:AP001507 PTS system, galactitol-specific enzyme II,  
 C component [Bacillus halodurans]  
 Identities = 92/347 (26%), Positives = 173/347 (49%), Gaps = 15/347 (4%)
- Query: 1 MKVTTGHLPIVDIGWQAGSLTAPSSBGLSPFFVGLLIRLGLFLGITRVFVPSNLWNN 60  
 MV G+ L ++D+GW A S A++S + GL++ + + + T+ + ++NN  
 15 Sbjct: 70 MVDRLGVDLNVLDVGVHPATSSIAWASVVAAPFIPLGLTVNVIMLVTKTKT-MNVDIWNF 128
- Query: 61 FGYMNGTMAYPATGNFILSPFAPMVFVLLYSVMSEVLADRMSEYYGVKNATINSINIE 120  
 + Y + Y + + I + V + +L +++ A SE+Y + +I + I  
 20 Sbjct: 129 WYTFMAAAYVTVSDSITIQALIAAVMFQIVALKVADWTAPMVSEFYELPGVSIATGSTIS 188
- Query: 121 TLIPALILDPLNNLGVNKKVKNPESLKTGLKGFCEMTLGFILGVITIGVLSRLNLSI 180  
 ++ + + G+ +P++++ + GIGE + +G ILG IG+L  
 Sbjct: 189 YAPGIMLVKGIQIKPGIKHKNADPDTIQRRPGTFGESIFGLILGAAIGLLAGYVNV--- 244
- Query: 161 DTWGGILGFAVALAAVMTIFPLITGVFASAPFLAEAVERNKKKESQABQALOKKRF 240  
 G ++ +A+AAVM + P + + P++E+ K + I  
 25 Sbjct: 245 ---GEVIRIGWMAAAVMVLMPEMVKILMBGLMPVSESARENKLR-----PDRRIHI 294
- Query: 241 AVDDGVGPGEPATIIAGLILVPMVVISLILPGNEALPVVDLIAIPFMIEAMIAVSKNI 300  
 +D V G P+ I LILVP+ V++++ILFQN LP DL IPF++ ++ ++GNT  
 30 Sbjct: 295 GLDAAVLLGRHPSVISTALILVLPLVLLAIVLPGNALLPFGDLATIFPFAFIVGAARGNI 354
- Query: 301 LKAILNGIWFPSLGLYAASALGPITYEAVKHFGTALPAGVTLMSFN 347  
 + ++L G I +L LY A+ + P++T+ ++ +P G LI S +  
 35 Sbjct: 355 IHSVLGAIMIALSLYMATDIAPVFTKMAENSFNFMPESSALISSID 401

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1500

- 40 A DNA sequence (GBSx1587) was identified in *S.agalactiae* <SEQ ID 4615> which encodes the amino acid sequence <SEQ ID 4616>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

- 45                   ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.1013(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 50 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1657-

**Example 1501**

A DNA sequence (GBSx1588) was identified in *S. agalactiae* <SEQ ID 4617> which encodes the amino acid sequence <SEQ ID 4618>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1294 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10147> which encodes amino acid sequence <SEQ ID 10148> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:RAC76604 GB:AB000435 L-xylulose kinase, cryptic [Escherichia coli K12]  
 Identities = 156/496 (31%), Positives = 261/496 (52%), Gaps = 18/496 (3%)

Query: 16 YYLSDYGGTNTKALIFDKLGHQIAVSSFETLKNETSGHGRQVNLVKTWNAITSAIRSVI 75  
 Y+L +D GG+ KA ++D+ G + V Q G + ++ + W + IR ++  
 Sbjct: 4 YNLGLDGGGSLKAGLYDRSGRAGVQRLPLCALSPQPMARMDMAELWQCCNAVIRALL 63

Query: 76 QISKLGFEQISAVACIGHGKGLNLLDNKLEPLEQGILSTONRAKDLAQYFESK--LDNIW 133  
 S +S RQI + GKGL+LLD +PL ILS+D RA ++ + ++ ++  
 Sbjct: 64 THSGVSGEQIVGIGISAQKGLPLLDKNDKPLGNAILSSDRAMEIVRRMQEDGIPKLY 123

Query: 134 ELTRQHIFPSQSPVILRWLKDQVPETYKSGIAGVLSAKDPTRYKLTGKVQOQYEGDASGNHW 193  
 LTRQ ++ +LRWLK+++PE Y IG V+ D++R+ LTR E + S ++  
 Sbjct: 124 PLTRQTLMTGHPVSLRWLKEHPERYAQIGCVNMTHDYLRWCLTGVEGCEESNISSENL 183

Query: 194 INFQITGYDPAILDFFGIRREIENSLPELIDGADLVPGOISSQAQKETGLVEGTFVVGGLF 253  
 N G YDP + D+ GI EI ++L+P ++ SA++ G I++Q A TGL GTPVVGGLF  
 Sbjct: 184 YNMSLGEYDPLTWNLGIAEINHALFPVVGSAKIC-GEITATGALTGLKAGTFVVGGLF 242

Query: 254 DIDACALGSGVLESQTFVSIGTWNINT--YPSLKPQKQSGIMTSYFFORRYLLEASSP 311  
 D+ + AL +G+ + T + + GTW ++ L+ + + Y D +++ +SP  
 Sbjct: 243 DVVSTALCAGIEDEFTLNAVGMTWAVTSGITRGLDGEARHPVYGRYVNDGEFIVHEASP 302

Query: 312 TSAQNLMFNLMLMHQIBDNKSSGSIYDNLLEFLHTDATHGLILFFPFLYGSNTSQD 371  
 TS+GNL + G +D + + + L F PFLYGSN +  
 Sbjct: 303 TSSGNLEMF-----TAQWGEISFDEINQAVASLPKAGGLFFLPFLYGSNAGLE 351

Query: 372 ASACFFGLTFKTSKQMRVAVYEGIAFAHKQHTIDLKSRGSPVKIIRFSGGATNSPAWM 431  
 ++ P+G+ T++ +++A+YEG+ P+H H+ + ++ R + +R +GG +S WM  
 Sbjct: 352 MTSGFYGMQALHTRAHLQAIYEGVVPFHMTHL-NRMRRFTVHTLRVTTGGPAHSDVM 410

Query: 432 QMFSDILNFPPIETVSGTELOGGLGAILARHALDKI-SLKEAVCDMVRKVAIKYKQLSEVK 490  
 QM +D+ IE + E G G A+ AR + EA +D+ P + + +  
 Sbjct: 411 QMLADVSGIRIKLPQVETGCGPAAALAAVRGTVGHYHNSFAQRLDRHPVRTLLPDMTARQ 470

Query: 491 GYKKKYHAYQKLESTL 506  
 Y+KKY YQ L+ L  
 Sbjct: 471 LYKKYQRYQHLLAAL 486

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1658-

**Example 1502**

A DNA sequence (GBSx1589) was identified in *S. agalactiae* <SEQ ID 4619> which encodes the amino acid sequence <SEQ ID 4620>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 59
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
10          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AG05648 GB:AE004652 hypothetical protein [Pseudomonas aeruginosa]
  Identities = 59/235 (25%), Positives = 104/235 (44%), Gaps = 9/235 (3%)

15  Query: 23  QVQLIKLVKDLGFSRPFETIRQELLQDFERELPAKKAENDFYDINLYYSANEELIK-GGKVN 81
      Q + L+ G R R B+R+EL P + AL A + +8+ +L+ G++N
      Sbjct: 23  QASFLPLLAAGAQRVKEELFAGEP-DTEALTAIQLQGLECVTFSSPLEKWRDQQLN 81

20  Query: 82  PYLNKGLKEASQLGAPPIKLNVGQTRNLKKELEPELKEIKSQTIGIKVZNQDPKATV 141
      P L L+ A GA ++K++G + +L L L + + +VEN+Q P+ +
      Sbjct: 82  PZLEPTLRARACGAGWLKVSIGLLPE--QPDLAALGRRLARHGLQLLVENDQTPQGRRI 139

25  Query: 142  ENQCYFHTLVKSLQIPISFVFDTANWAFINQDLYQAVNLAQDITTLHCKNFIQVAGKPH 201
      E + F L L+ Q+ ++ FD NW+ Q +A L Y+HCK I+
      Sbjct: 140  EVLERFPFLAERQQLDIAMTFDIGNWQCEQADEAALGRVIGVYHCKAVIRNEDGKL 199

30  Query: 202  LKSLFEGEINLTD-LKSFSPNCEYLALEYPT-LEILKRDVQRLISINSG 251
      ++ ++ LL+ F A+EYP + L +R+ + L + Q
      Sbjct: 200  VAVPPSADLIQVQRLQHPFSGVARAIEYPLQDDLLSLSPRHIAALRQGPQ 254

```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1503**

A DNA sequence (GBSx1590) was identified in *S. agalactiae* <SEQ ID 4621> which encodes the amino acid sequence <SEQ ID 4622>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 30
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
           bacterial cytoplasm --- Certainty=0.0430 (Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA03939 GB:AP001507 unknown conserved protein [Bacillus halodurans]
  Identities = 136/511 (26%), Positives = 234/511 (45%), Gaps = 29/511 (5%)

50  Query: 4  LDKKSYDLLFYLLKREBPETVMATANALNQSRKKVYYHLEKINDALPSDVQIVSYPRV- 62
      LD++S +L LL + + IN SRR VY LEKIN L + V R
      Sbjct: 3  LDKRSTFILLQLHARSYLPIDQLTKLVNSRRVYVNDGKINSWLEQGLKAGVYKVRSG 62

55  Query: 63  GILLTERQKAACRILLDEVTDSYVNMKSSERLQLSLVSIIVAKDRVTIDRMQLNDVSRN 122
      G++L E+ K L + + Y + ER ++ ++ + + + IM VSRN
      Sbjct: 63  GLILDERAKEBIFTKRLSLKSWIYBYSAQRKAWVVIYLLRLLEPLFLSHMDRGTVSRN 122

```

-1659-

Query: 123 TIANDELNELSELAKEYNLQLOSTKCRGYFLDGHL-----SIIQVLYKLDDIYHNGSS 178  
 T ++D+ L+ EL ++L L+ + GY + G +++ YL + L  
 Sbjct: 123 TTIDDIKCLKDEL--NNPHLALEFERKDGVTISGDTDKRKALVYLSQALPQQNWETKL 180

5 Query: 179 SPIDLFNHKLSQAGASTYPSKEVLVDYFHLYLFIQSRLGKKINSQQQFMQILPFIIM 238  
 S I + F L F + E L + + S++ L KI D L F+L  
 Sbjct: 181 SPIRIF---LTKRDNKGRIFTTELQKVDYVISESRKVL--KIQVTDVLESLSLRPLLF 235

10 Query: 239 AYRK-----MRLSPEVQTSLSNDFSLVWQKEYETAKLADELLENFQLSDELIEVGLVA 293  
 R +++ P + L KEYE AK ++ +L+ F + + EV +  
 Sbjct: 236 MKRVAKGRFKVHLEKQVLKGT-----KEYEAAKVMSPFLQAGVGHYPDEEVLYLT 288

Query: 294 MIMLSFRKDRIN-HLESQ-DYDMRATLTSFLKELEERYHLHFVHKKOLLRLTHCKAL 351  
 +LS + + N +ES+ + ++ +TS + + ++ + F K+ L + L H K  
 15 Sbjct: 269 THLSKINYANGELIESRKESQELTHIVTSMVNDQKYACVVFREKLEKILFPHIKPA 348

Query: 352 LYRKRYGIFSVNPLTEHIKDKYELFAITSSSVKLEKAWQTKLDDVAYLTIHLGSEL 411  
 YR +YG+ N + E IK Y EL+ T V LE+ + D++VA++T+H G +  
 Sbjct: 349 FYRKIGYLEVENNIASIKTSPYELFLTRKVVHYLERYGVKSNNEVAFITVHVGWM 408

20 Query: 412 RNSQGSFNK-LKLIVSDSGIAIQKLLKQCCRYLTSNDSIEAVPTBQYQSVSDLMHVM 470  
 R P K K +IV G+ + L Q + DI + +Y+ + VD  
 Sbjct: 409 RREGTIPTKRKKALIVCANGVTSQFLKNQLEGLFPVADIIKTCSIREYKTP--VEVDF 466

25 Query: 471 VVSTSALESRFPMLVNVHVPVITDDIIRLIR 501  
 ++ST+ E P +V+P+IL+ + RL++  
 Sbjct: 467 IISTSTIPEKNVPFIVNPIITETEKRELLK 497

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4623> which encodes the amino acid sequence <SEQ ID 4624>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.0745 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 368/548 (67%), Positives = 456/548 (83%)

Query: 1 MIILKKSYDLLFYLLKLEEFETVMAIANALNQSRKKVYHLEKINDALPSDVPQIVSYF 60  
 M+ILKKSYDLL YLAKLE PRVMAI++ALAQSRKKVY L+KIN ALP V QI+SYF  
 Sbjct: 1 MIILKKSYDLLSYLLKLETFETVMAISHALNQSRKKVYVQLKQALPKGVQDIISYF 60

45 Query: 61 RVGILLTEKQKAACRLLLEDVTDYSYVMKSSRLQLSLVSIIVAKDRVTIDRLMQNDVS 120  
 R+GILLT QKAACRLL+EVTDY+YVMK+ ER +LS + I V+ +RVTD+LQ+NDVS  
 Sbjct: 61 RLGITLADQKAACRLLLEEDVYHVMKSDERRRLSSITYAVSTERTVTDKLMQNDVS 120

50 Query: 121 RNTIANDLNLSELAKEYNLQLOSTKCRGYFLDGHPLSIIQVLYKLDDIYHNGSSSF 180  
 RNTIANDL ELR EL +K+Y +QL +TK RGY+ HP+++IQVLYKLL D+Y G++SF  
 Sbjct: 121 RNTIANDLTELKELRDKQYKQLHAYKARGYYFGCHPMALIQVLYKLLDVYQGSSSF 180

55 Query: 181 IDLFNHLKLSQAGASTYPSKEVLVDYFHLYLFIQSRLGKKINSQQQFMQILPFIIMAY 240  
 ID+FN KLS+ G S YRSK+L YFH YLF+SQ SLGK D+QD QFM+QLFPL+L+YF  
 Sbjct: 181 IDFNRLSEIQGLSVYFSKDLITYFHEYLFLSQASLGKTIINTQSQFMQILPFIIMAY 240

Query: 241 RMRKLSPEVQTSLSNDFSLVWQKEYETAKLADELLENFQLSDELIEVGLVAMMLSPR 300  
 R MRL E ++L +F L+H+RKEY LA++LA EL NP+L LD+YH +VAMMLSPR  
 60 Sbjct: 241 RMRKLSPEKALQSPHLEWRKRYHIAQLARELYHNFKLLDDIEVSNVAMMLSPR 300

Query: 301 KDRDNHLESQDYDMRATLTSFLKELEERYHLHFVHKKOLLRLTHCKALLYRKRYGIF 360  
 KD+D+H+ESQDYDMRAT++ P+ +LE RY LHF HK+DLL+L THCKAL+YRK YGIF  
 65 Sbjct: 301 KQDQHVESQDYDMRATISHPTDLQSRQYQLHPTKQDLKRLTHCKALLYRKRYGIF 360

-1660-

Query: 361 SVNPLTEHIKDKYEELFAITSSSVKLEKAWQIKLITDDVAYLTTHLQGLRLNSQQSPNK 420  
VNPLT+H+K+KYEELFA+T S +Lk+ W I LTDD+AYLTTHLQGLRL++ K  
Sbjct: 361 LVNPLTDHVKEKYELFAMTQSCATLEQQWTISLTDDIAYLTTHLQGLRLHNNTEQEK 420

5 Query: 421 LKLVIVSDGEGIAIQKLLKQCCRYLNSDIEAVFTTBQVQSVSDIMVDMVVSSTDALES 480  
KLVIVSD+GI IQKLL KQCCRYL N IEAVFTTBQVQSV DL+ VM+V+T+D L++  
Sbjct: 421 TKLVIVSDDGIGIQKLLFKQCCRYLANQITBAVFTTBQVQSVYDLLAVDMIVATDTLTK 480

10 Query: 481 RFLMLVAVHPVLITDDDIILIRLFRSKGNCNSNQFTNLEKTIQAQVKEDSERVVLAKSHK 540  
+ PML+V+P+L+DDDII+LIRFSK+G + ++P+ EL K I VK++S+RY L SKIE  
Sbjct: 481 KIFMLVAVHPILSDDDIILIRLFRSQGRKSEHSRSTELKALBAVVKEDSDRYALVSKIE 540

Query: 541 KLHQLLEL 548  
KLH+ELL  
15 Sbjct: 541 KLHRELEL 548

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1504

- 20 A DNA sequence (GBSx1591) was identified in *S. agalactiae* <SEQ ID 4625> which encodes the amino acid sequence <SEQ ID 4626>. Analysis of this protein sequence reveals the following:

Possible site: 50  
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2692 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77149 GB:AB000491 crf, hypothetical protein [Escherichia coli K12]  
Identities = 211/363 (58%), Positives = 270/363 (74%), Gaps = 9/363 (2%)

- 35 Query: 1 MPAVVDITRESWILSTFPENGWLNKEIEEVVARGNFAMWMLNGCGVMKCTPGCANVVM 60  
M VK ITRESWILSTFPENG+WLNEEIE+E VA G FAWMLG G+W+K+ CG NV +  
Sbjct: 3 MSKVKSTITRESWILSTFPENGSWLNERIEQEQVAPGTFAMWMLGCTGIWMLSEGGTVNVCV 62
- Query: 61 DLKSMRGKSTKKVDMVRGHQMMNAGVRLQPNLRAPQWVIDPFAINLIDYLVSHV 120  
D W GK + M +GHQM MGV+KLQPNLR P V+DPFAI ++D L +H H+  
40 Sbjct: 63 DFWOCTGKQSHGNPLMKQGHQMMNAGVRLQPNLRKTPFPVLDPFAIRQIDAVLATHDHN 122
- Query: 121 DHIDINTAAALNNINLNDHVKFGVPGYECGRHWKMGVFEERLIIVIKPGQESFEPKDIKVTA 180  
DHID+N AAL++N D V P+GP C ++W NGV+ER IV+KPG+ +KDI++ A  
45 Sbjct: 123 DHIDNVAAAVAMNC-ADDVPTIGPKTCVDMIGWGVFKERCIVVKGQVWVKDIETHA 181
- Query: 181 VESFDRTCIAPFLVDGAEHHDGLAGLAVIDEEMARKAVNNYIFTPGGTITHGADSHFSN 240  
+++FDR T L+TLE D + AG V + M +AVNY+F+TPGQ+YH DSH+SN  
50 Sbjct: 182 LDAFDRITALITLPADO-----KAAG--VLDPGMDDRAVNYLFTPGSLYHSGDSHYSN 233
- Query: 241 YFAKHGDKYKIDVAINNYGDNVPGIQDNTSIDLRMAENLRKAVIIPVHYDINSFMS 300  
Y+AKHG +++IDVA++YG+NP GI DNTS D+LR M L AKV+IP H+DINSNF A  
55 Sbjct: 234 YFAKHGNEHQIDVALGSGYGENPRGITDNTSADMLRMGEALNAKVIPPHIDINSNFPAD 293
- Query: 301 TDSILQLWQMRKERLQYDPHPFINEVGKQYTPQDKRIEYVHHPRGPDQCEQSHNIQFK 360  
EI LM+M+K+RL+Y F PRIN+VGK+T+P DKD KYH+PRGPDQCF E ++ FK  
Sbjct: 294 PQIRIVLWEMKDKRLKYGFKPTINVGKQYTPQDKRIEYVHHPRGPDQCEPTIEPLDPFK 353
- Query: 361 ALL 363  
+ L  
60 Sbjct: 354 SPL 356

-1661-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4627> which encodes the amino acid sequence <SEQ ID 4628>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3298 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 315/363 (86%), Positives = 348/363 (95%)

```

Query: 1  MPNVKIDITRESWILSTFPFWGTWLNKEIEHEVVAAGNFMMWMLGNGWIKITPGGANVVM 60
M  VADITRESWILSTFPFWGTWLNKEIEHEV  NFAMWMLGNGWIKITPGGANVVM
Sbjct: 1  MTKVQDITRESWILSTFPFWGTWLNKEIEHEVVPADNFAMWMLGNGWIKITPGGANVVM 60

Query: 61  DLMSNRGKSTKGVKDMVRGHQMANMAGVRLQPNLRAQPMVIDPFAINELDYLVSHPHS 120
DLMSNRGK-TK-VKDMVRGHQMANMAG  RLQPNLRAQPMVIDP  INELDYLVSH+HS
Sbjct: 61  DLMSNRGKATKGVKDMVRGHQMANMAGARKLQPNLRAQPMVIDPFAINELDYLVSHYHS 120

Query: 121 DHIDINTAAAIINRNPLDHVKVPVGPVECGEIVKMGVPEERLIIVIKPGSSPEFKDKVTA 180
DHIDINTAAAIINRN  L+RVKVPVGPVECGE+WK  MGVE++RI+++KPG+SPFKDKI+TA
Sbjct: 121 DHIDINTAAAIINRNKLNRVKVPVGPVECGEIVKMGVPEKRIIMILKPGDSSPEFKDKITA 180

Query: 181 VESFDRITCLVTLPVGDAGEHDGELAGLAVTDEMARKAVNYIFETPGGTYIHGADSHFSN 240
VESFDRITCLVTL+P+  DG+LAGLA+TD++MARKAVNYIFETPGGTYIHGADSHFSN
Sbjct: 181 VESFDRITCLVTLPIQGDAQDGLAGLAITDDMARKAVNYIFETPGGTYIHGADSHFSN 240

Query: 241 YFAKHGKDYKIDVAINNNGENPVGIGQDMTS-DLLRMAENLRKV+IPVHYDIMS NFAS 300
YFAKHG-DY IDV +NRYG+NP+GIGQDMTS-DLLRMAENLRKV+IPVHYDIMS NFAS
Sbjct: 241 YFAKHGRDYDIDVVLNNGENP+GIGQDMTSVLLRMAENLRKV+IPVHYDIMS NFAS 300

Query: 301 TDEILQWNRKERLQYDFHPFINEVGKKTYPQDKRIEYHHPRGPDCCFRQESNIQFK 360
TDEIL+LWNRKERLQYDFHPFINEVGKKTYPQD++RIEYHHPRGPDCC++SNIQFK
Sbjct: 301 TDEILELWNRKERLQYDFHPFINEVGKKTYPQDQNRIEYHHPRGPDCCFLDSNIQFK 360

Query: 361 ALL 363
ALL
Sbjct: 361 ALL 363

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1505

A DNA sequence (GBSx1592) was identified in *S.agalactiae* <SEQ ID 4629> which encodes the amino acid sequence <SEQ ID 4630>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3988 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10145> which encodes amino acid sequence <SEQ ID 10146> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1662-

&gt;GP:BAA18808 GB:D90917 hypothetical protein [Synecocystis sp.]

Identities = 358/785 (45%), Positives = 494/785 (62%), Gaps = 15/785 (1%)

- 5 Query: 22 LEKLDAMWRAANYISAQMYLKNFLRLRELVENDLKVHPIGHWGVPGNQFIYHLNRA 81  
L ++ +WRAANY++ +YL+DNPLLR L +K +GHNG+ DG +P+Y HLNR  
Sbjct: 44 LNMKGFWRAANYLAVGMVYLRNPLRLRELQPEQTEGRLHGWGSGFISPLVTHLNRI 103
- 10 Query: 82 INKYDILDMFYIEGPGHGGQVMVSNLYLDGYSYTLNHNIBOTEDGFKOLCKIFSPFGIAS 141  
I K+D DM Y+ GPGHG + YL+GSY+ + EDG K+ K PSPF GI S  
Sbjct: 104 IRKFPQDMLNMGVGGHGAAGPLGFCYLEGYSYRFAECSEEDRDMKRFFQSPFGIGIS 163
- 15 Query: 142 HAAPETPGSHBOGELGYALSHATGAILNDPNVDVIAVTVIGDGBGETGPIAGWLSNTTFIN 201  
H PETPGSHBOGELGY LSHA GA DNP++I + GDGE ETGFL W SN FIN  
Sbjct: 164 HCTPETPGSHBOGELGYCLSHAYGAADFNPNIIVGLAGDGESETGPLATSWHNSKFIN 223
- 20 Query: 202 PVMGDAVLPIFYLMGGKIHNPITFERKTDIELSOFFEGGLGWKPIFADVVELSEDAHAHA 261  
P+ DGAVLE+ +LNG KI+NP++ R+ EEL FEG G+ P P + D + H R  
Sbjct: 224 PIRGDAVLFVLHLNGYKINNPVLSRISREHLKALFEGGYTFYFVE---GSDPESMHQ 279
- 25 Query: 322 LPAEKLQAIQEIKTIOSEARQKPAEBAIQAKFPVLVARIPKGMTGPKAMBGTPIEGGFR 321  
A LD + BI IQ EAR A++ ++P+V R PKMTGTP +G +EG +R  
Sbjct: 280 AMAATLHCYSEIHQIQEASSTGI--AVRPWPMVNRTPKMTGTFYVGHKVEGWR 337
- 30 Query: 322 AHQVFPVDAHMHSHVDLSLWLGQSYRPEELFDENGKIVDEIAALSPKGRDRMSNPITN 361  
+HQVP+ + H+ L +W++SY+PEELFDE G + ALP+GD+R+ P N  
Sbjct: 338 SHQVPMGMHENPAHLQQLKAWMRSPKPEELFDEQGLKPGFALFEGKSLGSTPYAN 397
- 35 Query: 362 AGIV-KAMTADWKKFALDINVFQGINAQDMIEFGKYAADLVANDPNFRIGFDETKSN 440  
G++ ++ D+++ +D++ EG I A + G + D+ N NFR+PGFDE SN  
Sbjct: 398 GGLRLGLNMPDFRQYQIDVDQGTIBAPNTAPLGVLFDVMAWNNINFLPGFDENSN 457
- 40 Query: 441 RLQGVFTIRTSQWLGRRKPDYDEA--LSPAGRVIDSLSEHQAGFLSGYVLTGRHGFFA 498  
+L V+ ++ W+ + + LSP GRV++ LSEH EG+LE Y+L/GRHGFFA  
Sbjct: 458 KLVAVYEVSKKFWIAEYLEDQGLSGSPDGRVME-MLSEHTLGMLEAYLT/GRHGFFA 516
- 45 Query: 499 SYESFLRVVDSMVTQHFQWLRKSKTHITWRKNYPALNLIASSTVFQCDHNGYTHODPGIL 556  
+YESP V+ SMV QH KWL + H WR + +LN++ STV++QDHHG+THODPG L  
Sbjct: 517 TYESFAHVITSMVQHAQMLDLCR-HLNMRADISSLNILMTSTVMRQDHHGTHODPGIL 575
- 50 Query: 559 THLASKTPFYIREYLPADYNSLLAVMDKAPKAEDKINLIVTSKHPRPQFYSIAESELVA 618  
+ K+P+ +R YLP D NSLL+V D +++ IN+IV K Q+ + A  
Sbjct: 576 DVLINSPDVRVRIYLPFDVNSLLSVADHCLQSKYINIIIVCDQAHLYQDMTSAINRCT 635
- 55 Query: 619 EGYKVIDWASNVSLNQEPDVVFAAGTEPNLEALPAISILMKAFPELKIRFVNVDILKL 678  
+G + +WASN EPDVV AAG P EALAA ++L + FP LIRFV+V+D+LKL  
Sbjct: 636 KGVDIWASN-DAGTEPDVVMGAAGDIPTEKALATAMLRQFPFNLIRFVVDILKL 694
- 60 Query: 679 RHPSQDARGLSDESPNKVPTIDKVPVAFPHGYEDMIRDIFPSRHNH-NLH+GYRENGDI 737  
+ S+ GLSD +P+ +PTIDKPF+IF PH Y +I + + R NH NLH GY+E G+I  
Sbjct: 695 QPSEHPHGLSDRDFDLSFTDKPIIPNFHAYFWLHRLTYRSTHGNLHGYRENGDI 754
- Query: 738 TTPFDMRVNSELDRPHLAQDA--ALASLGNKAQAPSDENMMVAYHKDYIREHGDDIFPV 795  
TP D+ + ++DRF LA D L L + + M +Y EHG D+FE+  
Sbjct: 755 NTHMLAIQNDIRFSLAIDVIDRLPQIRVAGAHIKEMIKMDIGDGNAYENGIDMPRI 814
- 55 Query: 796 QNWK 800  
NN+W  
Sbjct: 815 VNRW 819

60 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.



**Example 1506**

A DNA sequence (GBSx1593) was identified in *S. agalactiae* <SEQ ID 4631> which encodes the amino acid sequence <SEQ ID 4632>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3509 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF37878 GB:AF234619 OpuAA [Lactococcus lactis]
Identities = 274/402 (68%), Positives = 338/402 (83%)

Query: 5 LEVKNLT/KIPGKKQAALMVKQKSKTEILEKTGATGVVDASFEIKGEIFVINGLSG 64
+++++LTKIPGK+ K AL M+ +G+ K EIL+KTGATGVVD +FEI EGEIFVINGLSG
Sbjct: 5 IKIEHLTKIPGRIKALTIVKEGEPKNEILKKTGATGVVDITNFEINGEIFVINGLSG 64

Query: 65 SGKSTLVRMLNRLIDPSSGNIYLDGKDIAQNVDELNIRRHIDIMVQNFGLFPFHITIL 124
SGKSTL+R+LNRLI+P+SG I+D +D+A +N EDL +RR ++MVQNFGLFPFHITIL
Sbjct: 65 SGKSTLLRLNRLIEPTSGKIFIDQDVATLNKEDLLQVRKTMVMVQNFGLFPFHITIL 124

Query: 125 ENTEFGLEMRGVSKKEERTTLAEKALDNLGLLPFKDQYPSOLSGMQQQRVGLARALANSPK 184
ENTE+GLE++ V KEER AEKALDNA LL FKDQY QLSGGMQQRVGLARALAN P+
Sbjct: 125 ENTEYGLGVQNPKEERKKRAEKALDNLNLLDFKDQYPSOLSGMQQQRVGLARALANPE 184

Query: 185 ILLMDEAFSALDPLIRREMDLELLDQDTNKQTIIFISHDLNEALRIGDRIALMDGQIM 244
ILLMDEAFSALDPLIRREMDLELLDQDTNKQTIIFISHDLNEALRIGDRIALMDGQIM
Sbjct: 185 ILLMDEAFSALDPLIRREMDLELLDQAKPQKTIIFVSHDLNEALRIGDRIALMDGQIM 244

Query: 245 QIGTGSEILTNPANDFVRFVVDVRSKVLTAQNMIMKPLTVLEIDGPQVALTRHREE 304
QIGTGSEILTNPAND+V+ FVRVDV+KV+TA+NIMI LTT +++DGP VAL +M EE
Sbjct: 245 QIGTGSEILTNPANDYVKTFFVVDVRAKVTIATNIMIPALTNIDVGSVALKMKTERE 304

Query: 305 VSMVMATNRRQLLGSLTADRAIRARKKDLPLSEVIDKDVTVTSKDTVITDIMPDIYDS 364
VS LMA +++RQ G +T++ AI ARK + PL +V+ DV TVSK+ ++ DI+P+IYD+
Sbjct: 305 VSSLMADVKKQFRGVTSEQAIARQINQPLKDVMTTQVTSKRMVLRDILPIYDAP 364

Query: 365 APIAVITENDRLGLVIRGVIRALANVQDETIVRSPKETVE 406
P+AV DEN L GV+IRG V+EAALA++ DE VE ++ E
Sbjct: 365 TPLAVVDNGLFGLVIRGVSLEALADIPDEDEVEEIEKREE 406

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4633> which encodes the amino acid sequence <SEQ ID 4634>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3761 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 344/395 (87%), Positives = 374/395 (94%)

Query: 1 MTNILEVKNLT/KIPGKKQAALMVKQKSKTEILEKTGATGVVDASFEIKGEIFVIM 60
M ILEVKNLT+KIPGKKQAALMVK GK+K+EI +KTGATGVVDASFE+K+GEIFVIM
Sbjct: 1 METILEVKNLSKIPGKKQAALMVKTKGNKSEIFKKTGATGVVDASFEVKKGEIFVIM 60

Query: 61 GLSGSGKSTLVRMLNRLIDPSSGNIYLDGKDIAQNVDELNIRRHIDIMVQNFGLFPFH 120

```

-1664-

GLSGSGKSTLVRLMLRLI+PS+G+I L+GKDI+ M+ + LR +RRHDINMVQF+P LFPF  
 5 Shjct: 61 GLSGSGKSTLVRLMLRLIEPSAGSILLBKDIDISTMSADQLREVRRHDINMVQSFALFPF 120  
 Query: 121 RTILENTEPGLMRGVSKERTTLAEKALDNAGLLPFDQYFSLQSGMQQRVGLARALA 180  
 +TILENTEPGLR+RGV KEER LAEKALDN+GLL FCDQYF+QLSGMQQRVGLARALA  
 Shjct: 121 RTILENTEPGLMRGVSKERTTLAEKALDNAGLLPFDQYFSLQSGMQQRVGLARALA 180  
 Query: 181 NSPKILMDAEPASALDPLIRREMQBELLDQDNTKQTIIFISHDLNEALRGDRIALMKD 240  
 NSPKILMDAEPASALDPLIRREMQBELLDQD+ KQTIIFISHDLNEALRGDRIALMKD  
 10 Shjct: 181 NSPKILMDAEPASALDPLIRREMQBELLDQDNTKQTIIFISHDLNEALRGDRIALMKD 240  
 Query: 241 GEIMQIGTGREILTNPANDFVRSFVSDVRSKVLTAQNMIMKFLITTVLSIDGPQVALTRM 300  
 G+IMQIGTGREILTNPANDFVRSFVSDVRSKVLTAQNMIMKFLIT +E+DGPQVAL RH  
 Shjct: 241 GQIMQIGTGREILTNPANDFVRSFVSDVRSKVLTAQNMIMKFLITTVLSIDGPQVALTRM 300  
 Query: 301 HREEVSMIMATNRRRQLGSLTADAATARKKDLPLSEVIDKVVTVSKDITVIDIMPLI 360  
 H EEVSMIMATNRRRQL+GSLTADAATARKK LPLSEVID+DV TVSKDT+ITDI+PLI  
 15 Shjct: 301 HNEEVSMIMATNRRRQLGSLTADAATARKKDLPLSEVIDRVRTVSKDITVIDIMPLI 360  
 Query: 361 YDSSAFIAYTDDNRLGVIIRGVIEALANVQDE 395  
 YDSSAFIAYTDDN+RLGVIIRGVIEALAN+ DE  
 Shjct: 361 YDSSAFIAYTDDNRLGVIIRGVIEALANISDE 395

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1507

A DNA sequence (GBSx1594) was identified in *S. agalactiae* <SEQ ID 4635> which encodes the amino acid sequence <SEQ ID 4636>. This protein is predicted to be OpuABC (opuAB). Analysis of this protein sequence reveals the following:

30 Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -10.67 Transmembrane 48 - 64 ( 43 - 72)  
 INTEGRAL Likelihood = -9.24 Transmembrane 101 - 117 ( 93 - 122)  
 35 INTEGRAL Likelihood = -7.54 Transmembrane 296 - 312 ( 290 - 316)  
 INTEGRAL Likelihood = -6.21 Transmembrane 252 - 268 ( 250 - 273)  
 INTEGRAL Likelihood = -5.57 Transmembrane 141 - 157 ( 138 - 170)  
 INTEGRAL Likelihood = -0.53 Transmembrane 220 - 236 ( 220 - 237)  
 ----- Final Results -----  
 40 bacterial membrane --- Certainty=0.5267 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GB:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]  
 Identities = 345/578 (59%), Positives = 429/578 (73%), Gaps = 8/578 (1%)  
 Query: 1 MENLQHLKLPVAPFVSTTNWITKTFGLWDFIQGNALMDWMTKTLFLNLPILPVL 60  
 M+L ++P+P+V+S T+M+T TFS FD IQ G L+L+T L+ L+I +  
 50 Shjct: 1 MIDLAICQVPPIANVSGATDWITSTFSGDFVIQKSGTVLWNCITGALTAVPFLMIAVV 60  
 Query: 61 TIAPVFLAKKWLQPLPTFTFGLLFTYQGLWQLNTFNVLNVLASLISIIIGVPLGIWMA 120  
 TI ++ KK P PTFGL I NQGLW L+T LVL+SL+SLIIGVPLGIWMA  
 55 Shjct: 61 TITALLVSGKKIAPPLFTFGLSLIANQGLMSDLMSITITVLLSLLSIIIGVPLGIWMA 120  
 Query: 121 KSDKVKQVNVNPLDFMQMDAPFVYLIPAVAFPGIGMVGVFASVVFALPPTVRPTNLIR 180  
 KSD V +V V PILDPMQMP FVYLIPAVAFPGIG+VPGVFASV+PALPPTVR TNL IR  
 Shjct: 121 KSDLVAKIVQPLDFMQMPFVYLIPAVAFPGIGVPGVFASVVFALPPTVRPTNLIR 180  
 60 Query: 181 EIPLELIEADSPSGSTVQKLFKVELPLAKNTIMAGINQTMALSNVTVGSMIGAPGLG 240  
 ++ EL+EA+DSFGST +QKLFK+E PLAK TIMAG+NOT+MIALSNVV SMIGAPGLG

-1665-

Sbjct: 181 QVSTELVEADSPGSTARQKLFKLEFPLAKGTIMAGVNTIMLASMVVIASMGAPGLG 240

Query: 241 REVLALQADIDGTGTVSGSLVLIAIVLIRVSPFNSKPGKEKQAKTSKVKK---VKLG 297  
R VLA+AQ ADIG GFVSG+SLVIAI++DR +Q N P EKQ + VKK + L

5 Sbjct: 241 RGVLAAGVQADIGKGFVSGISLVIAI+IDRFTQKIANVSPLEKQGNFT-VKKKRGIALV 299

Query: 298 ALALPILAAALCRIVVMTSGNEAKQKVKIANVQKDSVASTNVIAEVLKSKQYDVELTP 357  
+L I+ A M+ G A +KV +Y+ KDSVAS NV++ +K G+DV+ T

10 Sbjct: 300 SLIALIIGAFS-----GMSFGKTASDKKVLAVYMNNDSEVASINVLTQAMKEHGFVKTTA 355

Query: 358 LDNAVAMQTVANGADFTTSANLEKTHQGYFNKSLDDLGPHYVENVIGLAVPVMNV 417  
LDNAV MQTVANG AD SANLE TH + KY S+D LGR++ K+G VVP YNV

Sbjct: 356 LDNAVAMQTVANGQADGMSANLEPETHKTQMKYKGSVDLLGPNLKGAKVFPVVPSTNV 415

15 Query: 418 NSIEELNQADKQITGIEPGAGIMKSAQSLKDYVNLSSWKLISAFTGAMTTTLGKAIGN 477  
NSIE+L+NQR+K ITGIEPGAG+M +++++L Y NL WKL+ +9+GAMT LG+AIK

Sbjct: 416 NSIEDLTGANKTITGIEPGAGVMAASEKTLNLSYDNLKDWKLVPSSSGCMVALGEAIKQ 475

20 Query: 478 KDQVITGNSHWMFKYDLKYLKDPKKSFGSEHINTIARKNLKKMPKYYKIDKPKW 537  
+VITGNSHWMF KYDLKYL DK + G E+INTI RK LKK+ P YK+DKF T

Sbjct: 476 HKDIVITGNSHWMFNKYDLKYLKDPKGTMTSENTINTIVRKLKKENPRAYKVLDPKN 535

Query: 538 TKEDMESIMLMDKMEPAQAQKWKVNHKQVSEWTK 575  
T +ME+HLD+ G P +AA+ WK+HKEV +N K

25 Sbjct: 536 TTKMEAVMLDIQNGKTPFEAAKMWIKDHQKVDKWK 573

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4637> which encodes the amino acid sequence <SEQ ID 4638>. Analysis of this protein sequence reveals the following:

Possible site: 47

30 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood	Transmembrane	101 - 117 ( 93 - 121)
INTEGRAL	Likelihood = -7.54	Transmembrane	252 - 268 ( 250 - 273)
INTEGRAL	Likelihood = -6.85	Transmembrane	48 - 64 ( 43 - 70)
INTEGRAL	Likelihood = -5.57	Transmembrane	141 - 157 ( 138 - 170)
INTEGRAL	Likelihood = -5.26	Transmembrane	295 - 311 ( 289 - 315)
INTEGRAL	Likelihood = -0.53	Transmembrane	220 - 236 ( 220 - 237)

35

----- Final Results -----

40 bacterial membrane --- Certainty=0.4545(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]  
45 Identities = 340/571 (59%), Positives = 418/571 (72%), Gaps = 8/571 (1%)

Query: 8 KLPVAQVQLTQWIKTFSGGLFDINQVGSFLMDMHTKLLFIHPLFIVLVTAGNPL 67  
++P+ A V T+T+T TFS FD++Q G+ LM+ +T L + L I +T +

50 Sbjct: 8 QVPIANVSSATDITSTPSSGFDVIQKSGIVLMMGITGALTAVPFWIMIAVITLAILV 67

Query: 68 AKKKNPLPTFFLLGLLFTYQGLNKKQIMPTFLVNLASLVLGLIPLGINMKAKV 127  
+ KK P FT +GL I NQGLM LM+T TLVL++SL+S+IG+PLGINMK+ V +

Sbjct: 68 SGKKIAPPLPTFFIGLSLIANQGLMSDLASTITLVLLSLLSIITIGPLGINMKASDVL 127

55 Query: 128 IVNPILDFMQTPAPFYILIPAVAFPGICVPGVFASVIFALPPTVRFTHAIDLPTEL 187  
IV PILDFMQTPE FYILIPAVAFPGIC+VPGVFASVIFALPPTVR THL IR + TRH+

Sbjct: 128 IVQPIELDFMQTMPGYILIPAVAFPGICVPGVFASVIFALPPTVRFTHAIDLPTEL 187

60 Query: 188 EASDAPGSTARQKLFKLEFPLAKGTIMAGVNTIMLASMVVIASMGAPGLRGVIAAV 247  
EA+D+PGST +QKLFK+E PLAK TIMAGVNT+MLASMVV SMGAPGLRGV VI+A+

Sbjct: 188 EASDAPGSTARQKLFKLEFPLAKGTIMAGVNTIMLASMVVIASMGAPGLRGVIAAV 247

Query: 248 QHADIGSGFVSGSLVLIAIVLIRMTQLFNSKPGKEKQAKGKTNK---IGLAALAVFLIA 304  
Q ADIG GFVSG++LVIAI++DR TQ N P EK KW I L +L +I

65 Sbjct: 248 QSADIGKGFVSGISLVIAI+IDRFTQKIANVSPLEKQGNFTVKKKRGIAVSLIALIIG 307

-1666-

Query: 305 ALGRGIMAMTSGMADKGETVNIAYVQMDSEVASTHVAIEVLKNSGYHVTLTPLDNVAVWQ 364  
 A M+ G + V++ Y+ MDSEVAS +V+ + +K G+ V T LKNVAV WQ  
 5 Sbjct: 308 AFS----GMSFGKTASDKKVDLVYHNMDSSEVASINVLTKQAKHEHGDFVKTALDNVAVWQ 363

Query: 365 TVANGNADPSTSAWLPTVTHQQYQYKYSKLDDLGPNLKGTKLGLAVPKYMTDVNSIEDLS 424  
 TVANG AD SAWLP TH Q+QKY +D LGPNLKG K+G VP YM +VNSIEHL+  
 Sbjct: 364 TVANGQADGNVSAWLEINTHTKQYKYGKSVOLGPNLKGAKGVFVPSYH-NVNSIEDLT 422

Query: 425 KQADQKITGIEPGAGIMAAQKTLKEYENLSSWELVAASTGAMITSLDQAIKKDKPIVVT 484  
 QA++ ITGIEPGAG+MAA++KTL Y NL W+LV +S+GAMT +L +AIK+ IV+T  
 Sbjct: 423 NQAKNTITGIEPGAGVMAASEKTLNYSYONLKQMKLVPSSSGAMTVALGEAIKQHKDPIVT 482

Query: 485 AMSPHWMFAKYDLKYLKDKPEIFGSGTENINTIARKGLKKELPNVYKIIDKPHWTKQKMEA 544  
 WSPHWMF KYDLKYL DPK G++ENINTI RKGLKKE P YK++DKF+WT KQMEA  
 15 Sbjct: 483 GWSPHWMFNKYLKYLADPKNIGTSGTENINTIVRKGLKKENPEAYKVLDDKPNWTTDKMEA 542

Query: 545 VMLDINKMSPEAAAKKWKVANKSVSSWTK 575  
 VMLDI G +PE AAK W++ ++ +V W K  
 20 Sbjct: 543 VMLDIQNGKTPPEAAKNWIDHQKEVDKMPK 573

An alignment of the GAS and GBS proteins is shown below.

Identities = 439/576 (76%), Positives = 513/576 (88%), Gaps = 2/576 (0%)

25 Query: 1 MENLQHLKLPVAPPVESTINWITKTFSGLDFDIQTIGNALMDWMTKLLFINPLFLVILI 60  
 +E +LQ KLPVA VE T W+TKTSGLFD +Q +G+ LMDWMTKLLFI+PLLFIVLI+  
 Sbjct: 1 LETLITQLKPLVAQLVQGLTEWLTCTPSGLFDIMQVGSFLMDWMTKLLFIHPLFLFVLV 60

30 Query: 61 TIAVFLAKKKWQLPFTFTFGLGLFTYQGLMEQLINTFMVLVVASLISIIIGVPLGIWMA 120  
 T +FFLAKKKN LETFT +GLLFTYQGLM+QL+NTF LVLVASLIS++IG+PLGIWMA  
 Sbjct: 61 TAGMFPLAKKKWPLFTFTLGLLFTYQGLNQLMNTFTLVLVASLISVLIGIPLGIWMA 120

Query: 121 KSDKQKQVNVNPIIDFMQTMFAFVYLIPAVAFFGIGNVPGVPSVFALEPPTVRFTNLAIR 180  
 K+ V+Q+VNPILDFMQTMFAFVYLIPAVAFFGIGNVPGVPSVFALEPPTVRFTNLAIR  
 35 Sbjct: 121 KNATVQIVNPIIDFMQTMFAFVYLIPAVAFFGIGNVPGVPSVFALEPPTVRFTNLAIR 180

Query: 181 EIPLELEASD+FGSTVQKQLFKVELPLAKNTIMAGINQTMMLALSNVVTGSMIGAPGLG 240  
 +IP ELIASD+FGST KQKLFKVELPLAKNTIMAG+QTMMLALSNVVTGSMIGAPGLG  
 40 Sbjct: 181 DIPTELTRASD+FGSTGKQLFKVELPLAKNTIMAGVQTMMLALSNVVTGSMIGAPGLG 240

Query: 241 REVLSQLHADIGTGFGVSGLSIVLAIIVLDVRVQFNSKPGKEQAKTSIKVKQWGLGALA 300  
 REVLSQLHADIG+GFVSGL+LVLAIVLDR++Q FHSK EK AK K KW+GL ALA  
 Sbjct: 241 REVLSQLHADIGSGFVSGLALVLAIVLDRMTQLFHSKPGK-AKAGKNTKWGLGALA 299

45 Query: 301 LFLIALALGRIVNMISGNEAKGQKVIAYVQMDSEVASTHVAIEVLKNSGYHVTLTPLDN 360  
 +F++AALGR ++ MTSG K+G V IAYVQMDSEVASTHVAIEVLKNSGYHVTLTPLDN  
 Sbjct: 300 VFLIALALGRGIMAMTSGMADKGETVNIAYVQMDSEVASTHVAIEVLKNSGYHVTLTPLDN 359

50 Query: 361 AVNMQTVANGNADPSTSAWLPTTHGQ+ KYK+ LDDLSP++ +K+GL VPKYM +VNS 419  
 AVNMQTVANGNADP+TSAWLP THGQ+ KYK+ LDDLSP++ +K+GL VPKYM +VNS  
 Sbjct: 360 AVNMQTVANGNADPSTSAWLPTTHGQ+QYKYSKLDDLGPNLKGTKLGLAVPKYMTDVNS 419

Query: 420 IERLSNQADKQITGIEPGAGIMSAKQLKDYPLNLSWLLASTGAMTITLGAIKKHDK 479  
 IE+LS QAD++ITGIEPGAGIM +A++LK+Y NLSSW+L++ASTGAMT+L +AIK KD  
 55 Sbjct: 420 IEDLSKQADKQITGIEPGAGIMAAQKTLKRYHNLSWELVAASTGAMTITLQAIKKDK 479

Query: 480 QVITGWSPHWMFAKYDLKYLKDKPKSGPGEHINTIARKNLAKDMPKVKYKIIDKPHWTK 539  
 +V+T WSPHWMFAKYDLKYLKDKP+ FG E+INTIARK LK++P VYKIIDK PV+  
 Sbjct: 480 PIVVTWSPHWMFAKYDLKYLKDKPKSIFGSGTENINTIARKGLKKELPNVYKIIDKPHWTK 539

60 Query: 540 EDMESIMLMDKMEPAKAAQKWKIRKKEVSWTK 575  
 +DME++HLD++KGM P AA+KW++ +K +VS WTK  
 Sbjct: 540 KDMENVMLDINKMSPEAAAKKWKVANKSVSSWTK 575

65 A related GBS gene <SEQ ID 8827> and protein <SEQ ID 8828> were also identified. Analysis of this protein sequence reveals the following:



-1668-

```

|||: |:| ||| ||: ||| : ||||| ||| ||| : | |: ||| || ||| : |: ||: |||
DKLVPSSSGAMTVALGEATKQKHDIVITGWSFHWPFNKYDLYKLADPGTWTSTSENINTIVRKGLKGRNFEAYKVLDPK
470      480      490      500      510      520      530

5   1935      1965      1995      2025      2055      2085      2115      2145
KQVTEGMSINLMDKGMETAKAAQKQWIKRIRKGVSEMTK*YRKHVSFRACPLM*LKSP*LFNISPILF*YIKSERMKE
||: |||: |||: | | : ||: |||: |||: | |
NWTTKDEAVMLDIQNGKTFEERAAKNWIKQKRVKDFK
10 550      560      570

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1508

A DNA sequence (GBSx1596) was identified in *S. agalactiae* <SEQ ID 4639> which encodes the amino acid sequence <SEQ ID 4640>. This protein is predicted to be a transposase. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.65    Transmembrane    223 - 239 ( 223 - 240)
20  ----- Final Results -----
      bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10057> which encodes amino acid sequence <SEQ ID 10058> was also identified. A related GBS nucleic acid sequence <SEQ ID 10031> which encodes amino acid sequence <SEQ ID 10032> was also identified. A related GBS nucleic acid sequence <SEQ ID 10801> which encodes amino acid sequence <SEQ ID 10802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA50689 GB:X71844 putative transposase [Clostridium perfringens]
Identities = 94/364 (25%), Positives = 160/364 (43%), Gaps = 35/364 (9%)

35 Query: 8 KKHSLTLLDRNDIQSGELDRGETFKAIGLNLKHPTTIAKEVGN--KQLRESTKDLDCP 65
      K+KHL ++R ++ L G + L + T+ E+R KQ+++ +
      Sbjct: 12 KKHGLNKKRMVIEIRLKGDSATQNTYKELNRPINTVLNEIRGGTKQTKQSGKEPVVPA 71

40 Query: 66 LLKAPVYVNGCPKRRINGYKTFYLAQCAQRYVKLLVSRREGIPLNKSFTFWKIDRLV 125
      +A Y N ++N YK ++K +V+ K W D +
      Sbjct: 72 DTGEAVYKKN--RLKSNRYKLL-----ECSDFIKYVDKV-----KNDHMSLDACY 116

45 Query: 126 SNGVKKQRIYHILKTNDLVSSSTVRRHKKGLYSITPIDLPAVEFKKRAKSTLPPIP 185
      G+ ++ + +S+ T+Y ++ G L I IDLP K + +EST
      Sbjct: 117 -----GEALHSSRFSPSQIISTKTLYNVLDLGLPKINWDLF--AKLHNRKSTKVRNN 168

50 Query: 186 KAIKEGRYYEDFIHM-NQSEHNSLWENOTVIGRIGGK--VLTTFNVAPCNFIAKLMS 242
      K K G D +N+ E W E+D V+G K VLT + H G
      Sbjct: 169 KK-KIATSISDRNSTENREPEGHN-EIDCVIGKSKNDKVLITVVERKTYRAIIESSS 226

55 Query: 243 KTAIETAKHIQVTKRTLYDKRDFELPVLITDNGGEPARVEDIEDVCGSGQLFFCDP 302
      + I K + IK L F R+R I DNG RFA ++ E+ +++++P P
      Sbjct: 227 HSTISVTYKALDKKEPLGK---FSRVKSTITADNGSEFADLSEFLKT--KTKVTPTHP 281

Query: 303 NRSQDKARIEKHNTLVRLDPKGTSPNLTQEDINLALSHINSVKRQANCKTAYELPSP 362
      E +K E++ L+R +PKG ++E I+ ++N++ R+ L+ K+ ELF
      Sbjct: 282 YSSPENGNTNENHGLINRFTPKGRISDYSLETFISPIENWVTLPRKLLDYKTPEELPSI 341

Query: 363 TYGK 366

```

-1669-

K  
 Subjct: 342 HLEK 345

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1509

A DNA sequence (GBSx1597) was identified in *S.agalactiae* <SEQ ID 4641> which encodes the amino acid sequence <SEQ ID 4642>. Analysis of this protein sequence reveals the following:

10 Possible site: 33  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -11.30 Transmembrane 56 - 72 ( 48 - 79)  
 INTEGRAL Likelihood = -6.85 Transmembrane 11 - 27 ( 6 - 30)  
 15 INTEGRAL Likelihood = -6.69 Transmembrane 129 - 145 ( 126 - 158)  
 INTEGRAL Likelihood = -6.53 Transmembrane 94 - 110 ( 90 - 117)  
 INTEGRAL Likelihood = -1.54 Transmembrane 216 - 232 ( 215 - 232)  
 INTEGRAL Likelihood = -1.22 Transmembrane 147 - 163 ( 147 - 165)  
 ----- Final Results -----  
 20 bacterial membrane --- Certainty=0.5522 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 25 A related GBS nucleic acid sequence <SEQ ID 9431> which encodes amino acid sequence <SEQ ID 9432> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07666 GB:AP001520 unknown: conserved protein [Bacillus halodurans]  
 Identities = 112/224 (50%), Positives = 150/224 (66%), Gaps = 2/224 (0%)  
 30 Query: 8 IKDILWPIIFSLFGVLLMTFPKYNQMTTVAVSVISKTNQWNAVFPIHYIILLIIFIS 67  
 +KD LNP+IPS+ GV L M P + + T+ V+ ++K+ ++ P I+L I +  
 Subjct: 19 LKDYLANFLIPSIIIGVGLFWPIQKDNATITFVAFIAKOLQALDDHLFAILITMLAIV- 77  
 35 Query: 68 CVLALCYRLFRPSFIEKNDLLKEISDITIFMLIIRLIGLALGLMTVLHIGPEMWGKETG 127  
 VL+ LP+P+ KN LLK + I WL++R++G MT+L +GPE VW + TG  
 Subjct: 78 -VLSCVATLFPKPNLPMKGLLSLFIHPPMLVVRVLGFIAPMTLQLGPEAVWSEGTG 136  
 Query: 126 GLILFDLIGGLFTIPLAAGFILPFLTFEGLLRFVGVFLTFIMRFFPQLPGRSAVNCVASP 187  
 L+L+DL+ LFTIFL AG LPFL FGLLE GV L MRP F LPGRS+++C+AS+  
 40 Subjct: 137 ALLLYDLFLPLFTIPLFAGFLPFLINPGLLELFGVLINKPMRFVPTLPGRSSIDCLASW 196  
 Query: 188 VEDGTIGIALTDKQYVEGYTSREAAITSTTFSAVSITPCLXKL 231  
 +GDGTIG+ LT+KQY EG+YT RRAA ISTTFS VSITP + L  
 45 Subjct: 197 MGDGTIGVLTNKKQYEGFYTQREAAVISTTFSVVSITPISVVL 240

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1510

- 50 A DNA sequence (GBSx1599) was identified in *S.agalactiae* <SEQ ID 4643> which encodes the amino acid sequence <SEQ ID 4644>. This protein is predicted to be Na/H antiporter homolog (kefB). Analysis of this protein sequence reveals the following:

-1670-

Possible site: 17

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

5	INTEGRAL	Likelihood = -10.14	Transmembrane	176 - 192 ( 171 - 203)
	INTEGRAL	Likelihood = -9.34	Transmembrane	353 - 369 ( 348 - 373)
	INTEGRAL	Likelihood = -9.24	Transmembrane	3 - 19 ( 1 - 26)
	INTEGRAL	Likelihood = -7.17	Transmembrane	145 - 161 ( 142 - 168)
	INTEGRAL	Likelihood = -7.01	Transmembrane	86 - 102 ( 81 - 108)
10	INTEGRAL	Likelihood = -6.53	Transmembrane	52 - 68 ( 51 - 72)
	INTEGRAL	Likelihood = -5.79	Transmembrane	24 - 40 ( 23 - 49)
	INTEGRAL	Likelihood = -5.52	Transmembrane	214 - 230 ( 209 - 233)
	INTEGRAL	Likelihood = -4.04	Transmembrane	260 - 276 ( 258 - 278)
	INTEGRAL	Likelihood = -3.66	Transmembrane	287 - 303 ( 287 - 308)
15	INTEGRAL	Likelihood = -2.71	Transmembrane	113 - 129 ( 112 - 129)
	INTEGRAL	Likelihood = -2.66	Transmembrane	332 - 348 ( 330 - 349)

----- Final Results -----

	bacterial membrane	--- Certainty=0.5055 (Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000 (Not Clear) < succ>
20	bacterial cytoplasm	--- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:CAAS1756 GB:X73329 Na/H antiporter homologue [Lactococcus lactis]

Identities = 208/376 (55%), Positives = 285/376 (75%), Gaps = 3/376 (0%)

25	Query: 1	MHIIQITIIILLASVLATLISKRIIPAVVQQLVGIIIPAMLGVLHONQVHLVLSIG 60
	M+	I+Q+TI+L+AS++ATL S+R+ IPAV+Q+Q+LVGI+I P++LGLVH VL V+SEIG
	Sbjct: 1	MNDILQITIVLIASLIATLASRLIKPAVIGQMLVGILIPSVLGLVHSGHVLVMSIG 60
30	Query: 61	VILLMFLAGLEANFDLLKKYKPSLLVAITGVIVPNALFYPLRLRGPQINTAIFYGLVF 120
	VILLMFLAGLE++ +LKK K S+LVAI GVIVP+ +F + PG+ ++T+ FYG+VF	
	Sbjct: 61	VILLMFLAGLESDLTVLKKNFKSNLVAIGGVIVPLVIRGLVAFSGPYGMSTSFYFGIVF 120
35	Query: 121	AATISITIVVELQYKRVKKTDTGAIILGAADVAVLAVLLLSVFIA--TNGSSSNIGLQI 178
	AATG+SITVEVLQY +T T G+IILGAAD DD+LAVL+LS+F + GS ++V+	
	Sbjct: 121	AATSVSITIVVELQYKGLSTRAGSIILGAADVDDILAVLLLSIFTSFKNGSGSTHLFPQF 180
40	Query: 179	IIQLFFVFLFICMKYLVPALFKLTERVHFPEKYTILAILCPSLGLADKVGMSIIGS 238
	+++LFF FLF+ K L+P +K ++K+ K TI+A++IC LG+LAD VGMG++IGS	
	Sbjct: 181	ILLELFFAFILFVVHK-LIPRFWKFQKLPANKNTTIALIICLGLSLLDASVGMGSAVIGS 239
	Query: 239	FFAGLAIGQTSFVDKVBHKSILSGLYFFPIFFAGIALPLKFGDGMHMLTILITFALAV 298
	FFAGLAI QT K+E S +Y FIP+FF IA+ ++FD +H IL+FT Ia+	
	Sbjct: 240	FFAGLAISQTEVSHKIEYTSAGVVFIPVFVFLAISVQFSLIHPHWI ILFTLILAI 299
45	Query: 299	LSKLIPGYFVGRGNFSKLESITIGGMSVSRGEMALITVQGLAAKIISSTTYSELVIV 358
	L+K IP YFVG+ S ES+ IG GM+SRGEMALI+ Q+GL +II+ YSHIVIV+	
	Sbjct: 300	UTKPIPAYFVGSKNKLSTGSMILIGTGMISRGEMALIVAQIGTSAITTEDEVSELVIVI 359
50	Query: 359	ILSTIIAPPFILKYSPK 374
	IL+T++APF++K K	
	Sbjct: 360	ILATVLAFLIKLVK 375

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1511**

A DNA sequence (GBSx1600) was identified in *S.agalactiae* <SEQ ID 4645> which encodes the amino acid sequence <SEQ ID 4646>. Analysis of this protein sequence reveals the following:

Possible site: 22

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.



-1671-

## ----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14269 GB:Z99116 ypuA [Bacillus subtilis]  
 Identities = 86/319 (26%), Positives = 147/319 (45%), Gaps = 34/319 (10%)

10 Query: 3 IKLLFLAGLAFILFTLASPAYAASDVQKVIDETVYQPDYVIGSYLQKRAQTQLQLNYD 62  
 ++KK+ LA + L P + +D + + V LG L++ + + L +N  
 Sbjct: 1 MKKIWIGMLAAVLLLMVPKVSLADA--AVGDVIV---TLGADLSESDKQVLDENMPV 54

15 Query: 63 ESRDTVKRTINTSSYAKIMNADDAISQLY---SSVKIKKLSNDTLAVNIVTPENITK 118  
 ++ T V N + + +A I SS+ I K GS +N+ T NI+  
 Sbjct: 55 DNAIT-VTVINKEEHYLLKYSINQAIGSRAISSSSITIAKGGG---LNVET-HNIGS 108

20 Query: 119 VTEDMYRNAAVTLGIEHATISVAAPIKVTGESALAGIYYISLE-KNGASVSSNKQLAQHE 177  
 +T+MY NA +T G++ A + V AP +V+G +AL G+ + E + ++S + HQ+A +E  
 Sbjct: 105 ITDEYLNALMTAGVKDAKVVYTAPEVSGTAAALGLKAYEVSSDEALSEIDVQKQVANCQ 168

25 Query: 178 LSTLSGINAENKGEKGYDADKINVALTDIKSAVAKGSGSDLSKDDIRKIVETLKNYHLEN 237  
 L T S + + G E A + I K A K G + K D I K V++ + D+  
 Sbjct: 169 LVTISEL-GDKIGNENAAA-----LIAKIKKEFAKNGVFNKADIEKQVDDAASD--LNV 220

30 Query: 238 AVTENQINLIVNFAVNLQSQNVTKNSDFTNTLNNLKNIVSKAGSKFRKIVNVPNNKAV 297  
 +T++Q N+V S N +KN+D + D+ KA K + + +  
 Sbjct: 221 TLTDSQKNQLV-----SLFNMKKNADI--DWQVSDQL-DKARDKITFIESDEGKNFI 271

35 Query: 298 ESGKGFLANIWQQIVNFPQ 316  
 + F +IW IV+ F+  
 Sbjct: 272 QKVIDFFSVIWAIVSIFK 290

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1512**

A repeated DNA sequence (GBSx1602) was identified in *S.agalactiae* <SEQ ID 4647> which encodes the amino acid sequence <SEQ ID 4648>. Analysis of this protein sequence reveals the following:

40 Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.0603 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15719 GB:Z99122 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 76/138 (55%), Positives = 91/138 (65%), Gaps = 12/138 (8%)

50 Query: 1 MKLKAHEHIALIIVSDYRKSKDFYVKNLSPETIRENHRPERHDYKLDLC-GDIELEIFGN 59  
 M LK+EHIALI SDYRKSK FVY+KLSF++I+E +R ER YKLDL G +E+F  
 Sbjct: 1 MLKLSIHIALICSDYRKSKAFYVHKLGFPQVQISTYREERGSKYKLDLSNGSYVIELF-- 58

55 Query: 60 RLDDFPEYETPPQIGRPNWPREAQGLRHAFYVQDVEAYKVELNGLIFVEPIRYDDYT 119  
 + PP+R RP EA GLRHAF V ++ EL GI EPID TG  
 Sbjct: 59 -----SFPDPPPERQTRF-----EAAGLRHLAFTVGSLLKAVQELHEKGIETETPIRTOPLTG 109

-1672-

Query: 120 KMTFFFDPDGLPLELHE 137  
K+ TFFFDPD LPEL+E  
Sbjct: 110 KMTFFFDPDGLPLELHE 127

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4649> which encodes the amino acid sequence <SEQ ID 4650>. Analysis of this protein sequence reveals the following:

Possible site: 20  
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1205(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 99/137 (72%), Positives = 116/137 (84%)

- Query: 1 NKLAVHHIALIIVSDYKSKDFYVNNKLGFEIIRNHRPERHDYKDLRCGDIELEI PGNR 60  
NKL A+HH+ALIVSDY SKDFYVNNKLGFEIIRN+RP++HDYKDL CG IELEI P  
20 Sbjct: 2 MKNLAHHVALIVSDYHLSKDFYVNNKLGFEIIRNRYRPOKHDKYKJLSCGRLEI EIGKV 61  
Query: 61 LDDPEYETPPQRIQRPNMPRACGLRHFLAFVDPVEAYKVELENLIGFVPEIRYDDYTGK 120  
DP Y+ FP+R+ P+ EACGLRHFLAF V ++E+Y +L++LGI VEPTR+DDYTG+  
Sbjct: 62 TSDPNQAPPIRVSEPEFKSEACGLRHFLAFVTRNIESYVDDLKSLGIFVPEIRHDDYTG 121  
Query: 121 KMTFFFDPDGLPLELHE 137  
KMTFFFDPDGLPLELHE  
Sbjct: 122 KMTFFFDPDGLPLELHE 138

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1513

A DNA sequence (GBSx1603) was identified in *S.galactiae* <SEQ ID 4651> which encodes the amino acid sequence <SEQ ID 4652>. This protein is predicted to be alpha-amylase. Analysis of this protein sequence reveals the following:

Possible site: 40  
>>> Seems to have an uncleavable N-term signal seq  
INTEGRAL Likelihood = -11.62 Transmembrane 14 - 30 ( 7 - 36)

- 40 ----- Final Results -----  
bacterial membrane --- Certainty=0.5649(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AGG41778 GB:AF213261 sortase [Streptococcus gordonii]  
Identities = 136/247 (55%), Positives = 174/247 (70%), Gaps = 2/247 (0%)

- Query: 2 RNRKSHGPPFNFVNWLLVLLIIVGLALVNNKPIRNPFLAHQSNHYQIGRVSKKTEIENK 61  
R KK N + +L V+L+V LAL+FN IRN + +N YQ+S+VSKK IEINK  
50 Sbjct: 6 RRAKKGRSRRIILNLLSVLLIIVGLALVNNKPIRNPFLAHQSNHYQIGRVSKKTEIENK 65  
Query: 62 KSKTSYDPSVSKSISTESILSAQTKSHNLFPVIGGIAIPDVIEINLPFKGLNTELEVGAG 121  
SK S+P V+ +STG++L+AQ K+ LPVIGGIAIP++NLPIF GL H L YGAG  
55 Sbjct: 66 AKSGSFNFKEVPELSTEVANLQKWAQQLFPVIGGIAIPDLINLPFKGLNTELEVGAG 125  
Query: 122 TWKENQIMGPNYALASHRVFGLTGSKMLFSPLEHAKGKGVYLTDESKVYTYTITET 181  
TWKE Q M G NYALASHRVFG+TG+++MLFSP+ AK GMK+YLTDK KYTY+IT +

-1673-

Sbjct: 126 TMKETQEM-GKRYALASHHVPGITGANEMLFSPLDRAKGNKIYLTQKEKVYTTYSITSV 184

Query: 182 SKVTPFHEVIDD-TPGKSQLTIVTCTDPEATERIIVHAELEKTEGFSTADESILKAFSK 240  
V PE V+V+DD G +++TIVTCT D AT R IV LE + + IL F+K

Sbjct: 185 ENVEFERVDVDDAADGTAQVITATCRDAATSKTIVKGVLESETPYKSKPKILLNYPNK 244

Query: 241 KYNQIKL 247

YNQ+ L

Sbjct: 245 SYNQIKL 251

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4653> which encodes the amino acid sequence <SEQ ID 4654>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.12 Transmembrane 18 - 34 ( 13 - 38)

INTEGRAL Likelihood = -0.32 Transmembrane 94 - 110 ( 94 - 110)

----- Final Results -----

bacterial membrane --- Certainty=0.4248 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA73122 GB:M77279 alpha-amylase [unidentified cloning vector]

Identities = 60/122 (49%), Positives = 85/122 (69%)

Query: 7 RRKIKMSWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKQIKNNKEAKS 67

+ K + +W L+ +L I+GLAL+FN IR+ ++ +NS Y V+K+ +KNN ++

Sbjct: 4 KEKKGKGNLINSLLVLLPITIGLALIPNKGIRSWVVCQNSRSTAVSKLKPADVKNNMARET 64

Query: 68 TFDQAVEPVSSTESVLQAQMAAQLELVIGGIAIPELGINLPIFKGLNTELYGAGTKKEE 127

TFDF +VE +STE+V+AAQ + LPVIG IAP + INLPFKGL N L+ GAGTKKE+

Sbjct: 65 TFDPSVESLSTEAVMKAQFNNKLPVIGAIAPSVVEINLPFKGLSNVALITGAGTKMED 124

An alignment of the GAS and GBS proteins is shown below.

Identities = 147/245 (60%), Positives = 192/245 (78%)

Query: 2 RNNKSHGFFNFVRWLLVLLIIVGLALVFNKPIRNAPFAHQSNHYCISRVSKKTIKNN 61

+ K++ ++ R LL+ +L+I+GLAL+FNKPIRN IA SN YC+++VSKK I+KNN

Sbjct: 4 KQKRRKIKMSWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKQIKNN 63

Query: 62 KSTSYDPSSVKSISTESILSAQTKSHNLPVIGGIAIPDVKNLPIFKGLNTELYGAG 121

++K++DF +V+ +STES+L AQ + LPVIGGIAIP+ INLPFKGLNTELYGAG

Sbjct: 64 EAKSTFDQAVEPVSSTESVLQAQMAAQLELVIGGIAIPELGINLPIFKGLNTELYGAG 123

Query: 122 TMKENQIMGPNNYALASHHVPGITGSSQMLFSPLEHARKGNKIVLTDESKVYTTITTEI 181

TMKE Q+MOG NNY+LASHH+PG+TGSS+MLFSPLE A+ GM +YLTK K+Y Y I ++

Sbjct: 124 TMKEQVMQGHNNYSLASHHPIGITGSSQMLFSPLEHRAQNGNSIYLTDEKITYEIIKDV 183

Query: 182 SKVTPFHEVIDDTPGKSQLTIVTCTDPEATERIIVHAELEKTEGFSTADESILKAFSK 241

V PE V+VIDDT G ++TIVTCTD EATERIV EL+ +F A +LKA+

Sbjct: 184 FTVAPKRVVDIDTAGLKEVITVCTDIEATERIIVKGELEKTYEYDFDKAPADVLKAPNHS 243

Query: 242 YNQIN 246

YNQ++

Sbjct: 244 YNQVS 248

SEQ ID 4652 (GBS266) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 11; MW 26kDa).

GBS266-His was purified as shown in Figure 205, lane 10.

-1674-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1514

A DNA sequence (GBSx1604) was identified in *S. agalactiae* <SEQ ID 4655> which encodes the amino acid sequence <SEQ ID 4656>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1934 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4657> which encodes the amino acid sequence <SEQ ID 4658>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1934 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 711/819 (86%), Positives = 767/819 (92%)

Query: 1 MQDKNLVDNLTSEMKTSFIDYAMSVIVARALEVDVRLGKPVHRRILYGMNELGVTPDKP 60
      MQDKNLVDNLTSEMKTSFIDYAMSVIVARALEVDVRLGKPVHRRILYGMNELGVTPDKP
Sbjct: 1 MQDKNLVDNLTSEMKTSFIDYAMSVIVARALEVDVRLGKPVHRRILYGMNELGVTPDKP 60

Query: 61 HKKSARITGDVMGKYHPHGDSIYAMVRMAQWSYRHLVDGHNFGSMGDDGAAQRY 120
      HKKSARITGDVMGKYHPHGDSIYAMVRMAQWSYRHLVDGHNFGSMGDDGAAQRY
Sbjct: 61 HKKSARITGDVMGKYHPHGDSIYAMVRMAQWSYRHLVDGHNFGSMGDDGAAQRY 120

Query: 121 TEARMKSALEMLRLDINKNTVFDQNDYDGSEREPLVLPARFNNLVNGATGIAVGNATNI 180
      TEARMKSALEMLRLDINKNTVFDQNDYDGSEREPLVLPARFNNLVNGATGIAVGNATNI
Sbjct: 121 TEARMKSALEMLRLDINKNTVFDQNDYDGSEREPLVLPARFNNLVNGATGIAVGNATNI 180

Query: 181 PPHNLGESIDAVKLVMDNEDVITRELMEVIGPDPFTGALVMGRSGIHFRAYTGKGSIVL 240
      PPHNLGESIDAVKLVMDNEDVITRELMEVIGPDPFTGALVMGRSGIHFRAYTGKGSIVL
Sbjct: 181 PPHNLGESIDAVKLVMDNEDVITRELMEVIGPDPFTGALVMGRSGIHFRAYTGKGSIVL 240

Query: 241 RSRTEIETTSNGKERIVVTRFFYGVNKTQVHEHIVRLAQEKRLGKITAVDRSSREGVRP 300
      RSRTEIETTSNGKERIVVTRFFYGVNKTQVHEHIVRLAQEKRLGKITAVDRSSREGVRP
Sbjct: 241 RSRTEIETTSNGKERIVVTRFFYGVNKTQVHEHIVRLAQEKRLGKITAVDRSSREGVRP 300

Query: 301 VIEVRAASANVILNNLTKLSTQNTSPNMLAIEKGVPKILSLRQIIDNYIEHQKEVIV 360
      VIEVRAASANVILNNLTKLSTQNTSPNMLAIEKGVPKILSLRQIIDNYIEHQKEVIV
Sbjct: 301 VIEVRAASANVILNNLTKLSTQNTSPNMLAIEKGVPKILSLRQIIDNYIEHQKEVIV 360

Query: 361 RRTQPKAKAGARAHILEGLLVALDHLDESVITIIINSETDTIAQELMSRFELSERQSQ 420
      RRTQPKAKAGARAHILEGLLVALDHLDESVITIIINSETDTIAQELMSRFELSERQSQ
Sbjct: 361 RRTQPKAKAGARAHILEGLLVALDHLDESVITIIINSETDTIAQELMSRFELSERQSQ 420

Query: 421 ILDMRLRLTGLERDKIQSYNDLLALADLADILAKPERVVTIIKEEMDEVKKRYADAR 480
      ILDMRLRLTGLERDKIQSYNDLLALADLADILAKPERVVTIIKEEMDEVKKRYADAR
Sbjct: 421 ILDMRLRLTGLERDKIQSYNDLLALADLADILAKPERVVTIIKEEMDEVKKRYADAR 480

Query: 481 RTEIMGEVLSLEDEDLIEEDVILITLNNKGYIKRLAQDFRAQKRGGRGIGTGVNND 540
      RTEIMGEVLSLEDEDLIEEDVILITLNNKGYIKRLAQDFRAQKRGGRGIGTGVNND
```

-1675-

Sbjct: 481 RTBLAVGEVLSLEEDLKEEDVLITLSENKGYIKRLAQDEFAQRKGRGVQGTGVND 540  
 Query: 541 FVRELVSTSTHTDVLFPFNLGRVRLKAYEIPYGRTAGLPIVNLKLKLBSEGTIQTII 600  
 FVREL+STSTHTD+LFFPN GRVRLKAYEIPYGRTAGLPIVNLKLK+GETIQTII  
 Sbjct: 541 FVREL+STSTHTD+LFFPN+NGVRLKAYEIPYGRTAGLPIVNLKLK+EDGETIQTII 600  
 Query: 601 ARKE+VANKYFFFFTTQGGIVKRTSVSEFSNIRQNGLRALNKRDELINVLIDENEDVI 660  
 ARKE+ A K FFFTT+QSVIRKT VSEF+NIHQNGLR+ LKE D+LINVL +D+I  
 Sbjct: 601 ARKE+TAGSFFFTTKQGIKRTVESEFNNIRQNGLRALKRGLQNLINVLITSGQDII 660  
 Query: 661 IGTITGYSVRPKVNAVRRMKRTATGVGVNLREGDKVVGASRVNQGEVLITTEKGYGKR 720  
 IGT +GYSVRP ++RMR+ATGVGV LRE D+VVGASRI +GEVL+ITE G+GKR  
 Sbjct: 661 IGTISGYSVRPNEASIRMRGATGVGVNLREDDRVVGASRIQDQNGEVLITENGFGKR 720  
 Query: 721 TSASYPYTKRGKGKIKTANITAKNGPLARLAVITINGNEDIMVITDGVIIRTNVANISQT 780  
 T A++YPTKRGKGKIKTANIT KNG LA LVT++G EDIMVIT+ GVIIRTNVANISQT  
 Sbjct: 721 TSATDYPTKRGKGKIKTANITPKNQLAGLVTVDGTEDIMVITNKGVIIRTNVANISQT 780  
 Query: 781 GRSTMGVKMRDQAKIVTALVGEQIEDKSNIEDTKR 819  
 GR+T+GVK++LD +AKIVT LV+ E + I +E  
 Sbjct: 781 GRATIGVKIMKLDQAKIVTFLVQFEDSSIAEINTURE 819

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 1515

A DNA sequence (GBSx1605) was identified in *S. agalactiae* <SEQ ID 4659> which encodes the amino acid sequence <SEQ ID 4660>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA04010 GB:AJ000336 L-lactate dehydrogenase [Streptococcus pneumoniae]  
 Identities = 290/329 (88%), Positives = 313/329 (94%), Gaps = 1/329 (0%)

Query: 1 MTATQRKKVILVGDGAVGSSYAFALVNOGIAQLGIIIPALFDKAVGDAEDLSHALAF 60  
 MT+YKQKKVILVGDGAVGSSYAFALVNOGIAQLGIIIP L +KVGDA DLSHALAF  
 Sbjct: 1 MTSTKQKKVILVGDGAVGSSYAFALVNOGIAQLGIIIPQLHEKAVGDALDLSHALAF 60  
 Query: 61 TSPKKIYAATYADCADADLVITAGAPQKPGETRLDLVGKNLAINKSIIVTQVVSFNGI 120  
 TSPKKIYAA Y+DCADADLVITAGAPQKPGETRLDLVGKNLAINKSIIVTQVVSFNGI  
 Sbjct: 61 TSPKKIYAAQYSDCADADLVITAGAPQKPGETRLDLVGKNLAINKSIIVTQVVSFNGI 120  
 Query: 121 FLVAANPDVLTYSTWKPSGFKKRVIGSGTSLDSARFQALADKIGVDARSVHAYINGE 180  
 FLVAANPDVLTYSTWKPSGFKKRVIGSGTSLDSARFQAL+K+ VDARSVHAYINGE  
 Sbjct: 121 FLVAANPDVLTYSTWKPSGFKKRVIGSGTSLDSARFQALAEKLDVARSVHAYINGE 180  
 Query: 181 HGDSEFAVNSHANVAVQLBQWLQRNRDIDEGQLVDFISVRDAYSIINKKGAITYGIA 240  
 HGDSEFAVNSHAN+AGV LE++L++ +++ E L++LF VRDAY+IINKKGAITYGIA  
 Sbjct: 181 HGDSEFAVNSHANVAVQLBQWLQNVQEARLIELFEGVRDAYAITYIINKKGAITYGIA 240  
 Query: 241 VALARITKAILDDENAVLPLSVYQBGQYGVKVVFIQGPQAVGAGHIVRPVNIPLADAE 300  
 VALARITKAILDDENAVLPLSV+QBGQY V++VFIQGP+VGAHIVRPVNIPLADAE  
 Sbjct: 241 VALARITKAILDDENAVLPLSVYQBGQYGVNPFVFIQGPVGAHIVRPVNIPLADAE 300  
 Query: 301 QKQKASASQLKDIIDRAWNRNPFQKASKN 329  
 QKQKAS++L+ IIDRAWNRNPFQKASKN  
 Sbjct: 301 QKQKASAKELQALIDRAWNRNPFQKASKN 328

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4661> which encodes the amino acid sequence <SEQ ID 4662>. Analysis of this protein sequence reveals the following:

Possible site: 25

5

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.17 Transmembrane 106 - 122 ( 106 - 122)

10

----- Final Results -----

bacterial membrane --- Certainty=0.1468 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15

>GP:AB01558 GB:U60997 L(+)-lactate dehydrogenase [Streptococcus  
 bovis]  
 Identities = 278/329 (84%), Positives = 297/329 (89%), Gaps = 2/329 (0%)

20

Query: 1 MTATKQHKKVLVGDGAVGSSYAFALVNTQIAQELGIIDIFK--EKTQDAEDLSHALAF 58  
 MTATKQHKKVLVGDGAVGSSYAFALV Q IAEELGII+I + K GDAEDLSHALAF  
 Sbjct: 1 MTATKQHKKVLVGDGAVGSSYAFALVNTQIAQELGIIEIPLQFNKAVGDAEDLSHALAF 60

25

Query: 59 TSPKKIYAADYSDCHDADLVLTAGAPQKFGETRLDLVGNLAINKEVVTQIVASGFGKI 118  
 TSPKKIYAA Y DC DADLVV+TAGAPQKFGETRLDLV KNL INK +VT++V SGFGKI  
 Sbjct: 61 TSPKKIYAAYEDCADADLVITAGAPQKFGETRLDLVGNLAINKSIIVTESVVGSGFGKI 120

30

Query: 119 FLVAANFVDVLTYSYTWKFSGFFKERVIGSSTLSDSARFQALAAKIGVDARSVHAYINGE 178  
 FLVAANFVDVLTYSYTWKFSGFFKERVIGSSTLSDSARFQALAA K+ VDARSVHAYINGE  
 Sbjct: 121 FLVAANFVDVLTYSYTWKFSGFFKERVIGSSTLSDSARFQALAAKELVDARSVHAYINGE 180

35

Query: 179 HGDSEFAVNSHANVAGVGLYDNLQNRDIEBQGLVDFISVRDAAYSIINKKGATPYGIA 238  
 HGDSEFAVNSHANVAGV L +L+ +++E LV+LF VRDAYSIINKKGATPYGIA  
 Sbjct: 181 HGDSEFAVNSHANVAGVGLYDNLQNRDIEBQGLVDFISVRDAAYSIINKKGATPYGIA 240

40

Query: 239 VALARITKAILDDENAVLPLSVFQSQYBGVEDCYIQGPAIVGAGIVRPNVPIPLDAEL 298  
 VALARITKAILDDENAVLPLSVFQSQY V DCYIQGPAIVGA+GIVRPNVPIPLDAE  
 Sbjct: 241 VALARITKAILDDENAVLPLSVFQSQYANVTDCYIQGPAIVGAGIVRPNVPIPLDAEL 300

45

Query: 299 QIQMASANQLKAIIDEAFKKEEFASACKN 327  
 QIQ+ASA +LKAIDEAF+KEEFASA KN  
 Sbjct: 301 QIQMASAKELKAIIDEAFKKEEFASACKN 329

An alignment of the GAS and GBS proteins is shown below.

45

Identities = 286/329 (86%), Positives = 299/329 (89%), Gaps = 2/329 (0%)

50

Query: 1 MTATKQHKKVLVGDGAVGSSYAFALVNTQIAQELGIIDIFK--EKTQDAEDLSHALAF 60  
 MTATKQHKKVLVGDGAVGSSYAFALV Q IAEELGII+I +K GDAEDLSHALAF  
 Sbjct: 1 MTATKQHKKVLVGDGAVGSSYAFALVNTQIAQELGIIDIFK--EKTQDAEDLSHALAF 58

55

Query: 61 TSPKKIYAATYADCADADLVLTAGAPQKFGETRLDLVGNLAINKSIIVTQIVASGFGKI 120  
 TSPKKIYAA Y+DC DADLVV+TAGAPQKFGETRLDLV KNL INK +VTQ+V SGF GI  
 Sbjct: 59 TSPKKIYAADYSDCHDADLVLTAGAPQKFGETRLDLVGNLAINKEVVTQIVASGFGKI 118

60

Query: 121 FLVAANFVDVLTYSYTWKFSGFFKERVIGSSTLSDSARFQALAAKIGVDARSVHAYINGE 180  
 FLVAANFVDVLTYSYTWKFSGFFKERVIGSSTLSDSARFQALAA KIGVDARSVHAYINGE  
 Sbjct: 119 FLVAANFVDVLTYSYTWKFSGFFKERVIGSSTLSDSARFQALAAKIGVDARSVHAYINGE 178

Query: 181 HGDSEFAVNSHANVAGVGLYDNLQNRDIEBQGLVDFISVRDAAYSIINKKGATPYGIA 240  
 HGDSEFAVNSHANVAGV L WLQ NRDIEBQGLVDFISVRDAAYSI INKKGAT+YGIA  
 Sbjct: 179 HGDSEFAVNSHANVAGVGLYDNLQNRDIEBQGLVDFISVRDAAYSI INKKGATPYGIA 238

Query: 241 VALARITKAILDDENAVLPLSVFQSQYBGVEDCYIQGPAIVGAGIVRPNVPIPLDAEL 300  
 VALARITKAILDDENAVLPLSV+QSQY V+D +IQGPAIVGA+GIVRPNVPIPLDAEL  
 Sbjct: 239 VALARITKAILDDENAVLPLSVFQSQYBGVEDCYIQGPAIVGAGIVRPNVPIPLDAEL 298

-1677-

Query: 301 QIQMQASAEQLKIDIDEAWKNEPQKASKN 329  
 QIQMQASA QLK IIDEA+ EF A+KN  
 Sbjct: 299 QIQMQASAEQLKIDIDEAFAKEPFAKASKN 327

SEQ ID 4660 (GBS312) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 7; MW 40kDa).

GBS312-His was purified as shown in Figure 205, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1516

A DNA sequence (GBSx1606) was identified in *S. agalactiae* <SEQ ID 4663> which encodes the amino acid sequence <SEQ ID 4664>. This protein is predicted to be NADH oxidase (nox). Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial: cytoplasm --- Certainty=0.1888 (Affirmative) < succ>  
 bacterial: membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial: outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA26485 GB:AF014458 NADH oxidase [Streptococcus pneumoniae]  
 (ver 2)  
 Identities = 363/458 (79%), Positives = 408/458 (89%), Gaps = 3/458 (0%)  
 Query: 1 MSKIVVVGINHAAGTAITMLNLYGEANEIVFPDQNSNISFLGCGMALWIGBQIDGPEGL 60  
 MSKIVVVG NHAAGTA I TML N+G NEIV FPDQNSNISFLGCGMALWIGBQIDG KGL  
 Sbjct: 1 MSKIVVVGANHAAGTACINTMLNFGNENEIVFPDQNSNISFLGCGMALWIGBQIDGAGGL 60  
 Query: 61 FYSDEKEQLSNGAKVYMSPVLSIDYDKKEVTALVDGKEHVESYEKLIATGSGPIIPI 120  
 FYSDEK+LE+ GARVYMSPVLS+IDVD K VTA V+GKHH ESEYKLI ATGS PI+PIPI  
 Sbjct: 61 FYSDEKELEAGAKVYMSPVLSIDYDNKVVTAASVSGKEHKBSEYKLIPIATGSGTPIIPI 120  
 Query: 121 KGVIEIQGSRREPATLENLQFVKLYQNSEEVIEKLAQK--INRVAVVGAGYIGVELAEA 178  
 +GVEI +G+REPATLEN+QFVKLYQN+EEVI KL+ ++R+AVVG GYIGVELAEA  
 Sbjct: 121 EGVEIVKGNREPATLENVQFVKLYQNAREVINKLSDKSHLDRIAVVGSGYIGVELAEA 180  
 Query: 179 FORIGKEVTLVDVADTCMGYVDRDPTDMNSQNLEDHGIIRLAFQQAQVAVGSGKVERLV 238  
 P+R+GKGV LVD+ DT+ GYVD+DPT NM+KNLEH IRLA GQ V+A+EGDGKVERL+  
 Sbjct: 181 PERLGKEVTLVDVDTVLNGYVDKDPQNMKNLEHDIRLALGQVKAIRGDKGKVERL 240  
 Query: 239 TDKETFDVVMVILAVGFRPNTLGGKGLDTPRNGAVVVDKKQETSVDVDVYAIQDCATWD 298  
 TDKE+FDVVMVILAVGFRPNT L GK++ PRNGA+VDDKQETS+ VYA+QDCAT++D  
 Sbjct: 241 TDKESFDVVMVILAVGFRPNTALDGKLEPRNGAFLVDKKQETSIPGVYAVGDCATVD 300  
 Query: 299 NSRDDNYIALASNAVRTGIVAAHNACTELGAGVQSGNGISITGLNWSVTGLTLEKAK 358  
 N+R D +YIALASNAVRTGIV A+NACQ ELGG VQSGNGISITGLNWSVTGLTLEKAK  
 Sbjct: 301 NARKDTSYIALASNAVRTGIVGAYNACHELEGIGVQSGNGISITGLNWSVTGLTLEKAK 360  
 Query: 359 GAGYNAVETGFNDLQKPEPIKHNNEHVAIKIVDKDSRVILGQMVSHR-DVSGIHMPS 417  
 AGYNA ETGFNDLQKPEF+KH+NHEVAIKIV+DKDSR ILG QMVSH+ +SGIHMPS  
 Sbjct: 361 AAGYNATETGFNDLQKPEPKHNDHEVAIKIVDKDSRILGQMVSHDIASGIHMPS 420  
 Query: 418 LAIQEKVTIEKALTDIFPLPHFNKPNYNYITMAALGAK 455  
 LAIQE VTI+KIALTD+FFLPHFNKPNYNYITMAAL A+  
 Sbjct: 421 LAIQEVTIDKIALTDLFFLPHFNKPNYNYITMAALTA 458

-1678-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4665> which encodes the amino acid sequence <SEQ ID 4666>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2068 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 362/456 (79%), Positives = 403/456 (87%)

Query: 1 MSKIVVVGTHNGTAAIKIMLSNYGEANEIVFDQNSNISFLGCGMALWIGEQTIDGPEGL 60  
 MSKIVVVG NHAGTA IKIML+NYG+ANBIV FDQNSNISFLGCGMALWIGEQT GPEGL  
 Sbjct: 1 MSKIVVVGANHAGTACIKIMLINYGDANSIVVFDQNSNISFLGCGMALWIGEQTAGPEGL 60

Query: 61 FYSDEQLQESMGAKVYMSFVLNIDYDKKEVTALVDGKHEVSEYKLLATGSQPIIFPI 120  
 FYSDE+LES+GAKVYM SFV +IDYD K VIALVDGK HVE+Y+KLI ATGSQPI+PFI  
 Sbjct: 61 FYSDEKELESGLAKVYMSFVQSIDYDAKTALVDGKNHVETYKLI FATGSQPIIFPI 120

Query: 121 KGVEIQBGSREFKATLENLQFKLYQNSSEEVIEKLAKPGINRVAVVGAGYIGVELAEAFQ 180  
 KG EI+BSG KF+ATLENLQFKLYQNS +VI KL I RVAVVGAGYIGVELAEAFQ  
 Sbjct: 121 KGAEIKEGSLEFEATLENLQFKLYQNSADVIKLENKDIKRVAVVGAGYIGVELAEAFQ 180

Query: 181 RIGKEVILVDVADTCMGSYDRDFDTMMSKNLEDHGGIRLAFGQAVQVBDGKVERLIVTD 240  
 R GKEV L+DV DTC+ GYDRED TD+M+KN+E+HGI+LAFG+ V+ G+GKVE+I+TD  
 Sbjct: 181 RKGKEVILVDVDTCLAGYDREDITDLMAKNMEEHGQILAFGETVKEVANGKVEIITD 240

Query: 241 KETFDVDMVILAVGFRPNTLGGAGKLDITFRNGAVVDKKQETSVDVYAIGDCATTWNS 300  
 K +DVMVILAVGFRPT LG GK+D FRNG+V+K+QETS+ VYAIGDCATT+DN+  
 Sbjct: 241 KNEYVDVMVILAVGFRPNTLGGNKIDLFRNGAFLVNRKQETSIFGVAIGDCATTYDNA 300

Query: 301 RDDINYLALASNAVRTGIVAAHNACTLEGAGVQSGNSIISYGLNMVSTGLTLEKARQA 360  
 D NYIALASNAVRTGIVAAHNACT+LEG GVQSGNSIISYGL+HVSSTGLTLEKAR+  
 Sbjct: 301 TRDINYLALASNAVRTGIVAAHNACTDLEGIGVQSGNSIISYGLHVSSTGLTLEKARQL 360

Query: 361 GYNVETGPNLDQKPEFIKRGNEVAIKIVYDKDSRVILGCGMVSHEDEVSMGIRHPSLAI 420  
 G+H+ T + D QKPEFI+H N V IKIVYDKDSR ILG QM + EDVSMGIRHPSLAI  
 Sbjct: 361 GFDRAVETGPNLDQKPEFIEHGNFVTIKIVYDKDSRRLGAGQAAREDEVSMGIRHPSLAI 420

Query: 421 QESVTIEKLALTDIFFLPHFNKPNYITMAALGAKD 456  
 CE VTIEKLALTDIFFLPHFNKPNYITMAALGAKD  
 Sbjct: 421 QESVTIEKLALTDIFFLPHFNKPNYITMAALGAKD 456

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1517

A DNA sequence (GBSx1607) was identified in *S.agalactiae* <SEQ ID 4667> which encodes the amino acid sequence <SEQ ID 4668>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2319 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.



-1679-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1518

- 5 A DNA sequence (GBSx1608) was identified in *S.agalactiae* <SEQ ID 4669> which encodes the amino acid sequence <SEQ ID 4670>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
10  INTEGRAL    Likelihood = -7.75    Transmembrane  160 - 176 ( 157 - 179)
    INTEGRAL    Likelihood = -7.38    Transmembrane   73 - 89 ( 70 - 97)
    INTEGRAL    Likelihood = -5.47    Transmembrane  289 - 305 ( 284 - 312)
    INTEGRAL    Likelihood = -4.09    Transmembrane  107 - 123 ( 106 - 124)
    INTEGRAL    Likelihood = -3.24    Transmembrane   43 - 59 ( 43 - 59)
15  INTEGRAL    Likelihood = -1.91    Transmembrane  258 - 274 ( 258 - 275)
    INTEGRAL    Likelihood = -1.33    Transmembrane  234 - 250 ( 233 - 251)
    INTEGRAL    Likelihood = -0.00    Transmembrane  209 - 225 ( 209 - 225)

----- Final Results -----
20  bacterial membrane --- Certainty=0.4100 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9805> which encodes amino acid sequence <SEQ ID 9806> was also identified.

- 25 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
Identities = 172/318 (54%), Positives = 234/318 (73%)

30  Query: 5  LSLTTIFALLPSSMLIYATPLIFTSIGTTFSEGGIVVNGLEGIMVIGAFSGVVFNLEFA 64
    + + I + + + + L+YA FLI T++GG FSEER G+YH+GLEG+M+IGAF+ V+ENL F
    Sbjct: 1  MDIVQILSLIIVPATLVYAAPLILTALGSGVFSERSGVVNGLEGIMIGAFSTVLFLNLPFG 60

    Query: 65  SVPGDATFWISVLGGLVGLIFSVIRHATVNFRAHDIISGTVNLMAFSLAVFLIKLVY 124
    G A FW+S+L G +FS+IHA A ++FRAD +SG +N+AA +F++K++Y
35  Sbjct: 61  QELGAAPWLSLLAAMAGALFSLIHAAAIISFRADQTVSGVAINMALGATFLIVKLTY 120

    Query: 125  NKQQTDMIQESFGKFNPFILSDIPFGDIFPKGTSLVGYIALFLSFANFLYKTRPGLR 184
    K QID I E F K P L DIP +G IFF +AI +F++WFIL+KT FGRL
40  Sbjct: 121  GKQATDKITPEFPYKTKIPLGLGDIPVLGKIFFSDVYVTSILALALAPISWFLPKTPFGLR 180

    Query: 185  IRLSVGEHPQAAATLGINVYLMRYSGVLISGFLGGIGCAVYAGSISVNFMAATTLIGRGPIS 244
    +RLSVGEHP AADT+GINVY MEY GV+ISG GG+GG VYA +I++F +TI G GPI+
45  Sbjct: 181  IRLSVGEHPQADTMGINVYLMRYIGVMSGLFGLGGGVYASTLLDPTHTISTISGQFLA 240

    Query: 245  LAAMIFGKNQPIQRMILASLFFGLSQSLAVIGSHLPLLSNIPVTYLIAPVVLITIVLAAP 304
    LA++F+GKM+PIGA+ A+LFFS +QSL++IGS LPL +ID VY+ +AFY+LTI+ L F
50  Sbjct: 241  LAALVFGKWHPIGALGNALFFGFAQSLISLIGSLPLFKDIPWVYMLMAFYILTILALGPF 300

    Query: 305  PGQAVAPKADGINYIKTK 322
    G+A APK+G+ YIK K
    Sbjct: 301  IGRADAPKANGVPYIKGK 318

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4671> which encodes the amino acid sequence <SEQ ID 4672>. Analysis of this protein sequence reveals the following:

```

55  Possible site: 22
    >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -8.92    Transmembrane  73 - 89 ( 69 - 97)

```

-1680-

5	INTEGRAL	Likelihood = -5.04	Transmembrane	160 - 176 ( 158 - 177)
	INTEGRAL	Likelihood = -4.62	Transmembrane	289 - 305 ( 284 - 312)
	INTEGRAL	Likelihood = -3.98	Transmembrane	234 - 250 ( 232 - 251)
	INTEGRAL	Likelihood = -2.13	Transmembrane	107 - 123 ( 106 - 123)
10	INTEGRAL	Likelihood = -2.02	Transmembrane	43 - 59 ( 43 - 59)
	INTEGRAL	Likelihood = -0.53	Transmembrane	258 - 274 ( 258 - 274)
----- Final Results -----				
10	bacterial membrane		--- Certainty=0.4567 (Affirmative)	< succ>
	bacterial outside		--- Certainty=0.0000 (Not Clear)	< succ>
	bacterial cytoplasm		--- Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

15	>GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]			
	Identities = 176/318 (55%), Positives = 239/318 (74%)			
20	Query: 5	MSLVTFPALLMSMLIYATPLIFTSIGTFSERSGVVNVGLSGIMVWGAFSQIVTLEFA	64	
	Sbjct: 1	M +V I + + + + L +YA PLI T + +GG FSESGVVN+GLEG+M+ +GAF+ ++PNI F		
25	Query: 65	ETFGKATPWIAVLVGGIVGLIFSLIHAVATINFRADHIVSGTVNLNLAFAVFLVKMY	124	
	Sbjct: 61	+ G A PW++L G +PSLIHA A I+FRAD VSG +N+LA +F+VK +Y		
30	Query: 125	GKGQTDNIQSQSGPKGDFPGLSQIPVIGDIPFKTSLIGYTAIAPSPFANFLLYKTRGLR	184	
	Sbjct: 121	GK QTD I + F K PGI IPV+G IFF + AIA +F +WF+L+KT PGLR		
35	Query: 185	IRSVGSEHPQAADTLGINVYLMKYGVMSGLFGIGGAVYAQSISVNFVTTILGPGFIA	244	
	Sbjct: 181	+REVGEHP AADT+GINVY M+Y GVMISG GQ+GG VYA +I+++F +TI G GFIA		
40	Query: 245	LAAMIPGKNWPGAMLSLFFGLSQSLAVIQAQLPLEIKPTVYLQIAPVMTIILAAF	304	
	Sbjct: 241	LAA++PGKN+P+GA+ ++LFFG +QSL++IG+ LEL + IP VY+ +APY++TI+ L F		
45	Query: 305	FGQAVAPKADGINYIKSK 322		
	Sbjct: 301	G+A APKA+G+ YIK K		

An alignment of the GAS and GBS proteins is shown below.

Identities = 272/322 (84%), Positives = 301/322 (93%)				
45	Query: 1	MYSKLSLTTIFALLFSSMLIYATPLIFTSIGTFSERGVVNVGLSGIMVWGAFSQIVTLEFA	60	
	Sbjct: 1	+V+K+SL TIFALL SSMIYATPLIFTSIGTFSER G+VNVGLSGIMV+GAFSG+VFN		
50	Query: 61	LEFAVFGDATPWISVLVGLVGLIPSVIHAVATNFRADHISGTVNLNLAFAVFLV	120	
	Sbjct: 61	LEFA PG ATW+L+VWVG+VGLIFS+IHAVAT+NFRADHI+SGTVNLN+APS AVFL+		
55	Query: 121	KVLYNKQQTNIQSQSGPKGNPFILSDIPVIGDIPFKTSLVGYTAIAPSPFANFLLYKTR	180	
	Sbjct: 121	K+Y KQQTNIQ+SGPKG+FP LS IP +GDIFPK TSL+GY AI PSF AMF+LYKTR		
60	Query: 181	FGRLRSVGEHPQAADTLGINVYLMKYGVMSGLFGIGGAVYAQSISVNFVTTILGP	240	
	Sbjct: 181	FGRLRSVGEHPQAADTLGINVYLM+Y GV+ISGPIGGIGGAVYAQSISVNFVTTILGP		
65	Query: 241	GFISLAAMIPGKNWPGAMLSLFFGLSQSLAVIQAQLPLEIKPTVYLQIAPVMTIIL	300	
	Sbjct: 241	GFIA+LAAMIPGKNWPGAMLSLFFGLSQSLAVIG+ LPLL IPTVYLQIAPVMTIIL		
70	Query: 301	LAAPFGQAVAPKADGINYIKSK 322		
	Sbjct: 301	LAAPFGQAVAPKADGINYIK+K		



-1682-

290 300 310

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 5 Example 1519

A DNA sequence (GBSx1609) was identified in *S.galactiae* <SEQ ID 4673> which encodes the amino acid sequence <SEQ ID 4674>. This protein is predicted to be ribose/galactose ABC transporter, permease protein (rbsC-1). Analysis of this protein sequence reveals the following:

Possible site: 55  
 10 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -14.59 Transmembrane 205 - 221 ( 200 - 228)  
 INTEGRAL Likelihood = -13.69 Transmembrane 21 - 37 ( 13 - 45)  
 INTEGRAL Likelihood = -7.27 Transmembrane 302 - 318 ( 290 - 321)  
 15 INTEGRAL Likelihood = -7.17 Transmembrane 115 - 131 ( 111 - 138)  
 INTEGRAL Likelihood = -4.25 Transmembrane 251 - 267 ( 250 - 268)  
 INTEGRAL Likelihood = -2.97 Transmembrane 63 - 79 ( 63 - 80)  
 INTEGRAL Likelihood = -2.87 Transmembrane 333 - 349 ( 328 - 349)  
 ----- Final Results -----  
 20 bacterial membrane --- Certainty=0.6838(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8831> which encodes amino acid sequence <SEQ ID 8832>

25 was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6  
 SRCFLG: 0  
 MoG: Length of UR: 24  
 Peak Value of UR: 3.06  
 30 Net Charge of CR: 3  
 MoG: Discrim Score: 12.53  
 GvH: Signal Score (-7.5): -5.31  
 Possible site: 46  
 >>> Seems to have an uncleavable N-term signal seq  
 35 Amino Acid Composition: calculated from 1  
 ALOM program count: 7 value: -14.59 threshold: 0.0  
 INTEGRAL Likelihood = -14.59 Transmembrane 196 - 212 ( 191 - 219)  
 INTEGRAL Likelihood = -13.69 Transmembrane 12 - 28 ( 4 - 36)  
 INTEGRAL Likelihood = -7.27 Transmembrane 293 - 309 ( 281 - 312)  
 40 INTEGRAL Likelihood = -7.17 Transmembrane 106 - 122 ( 102 - 129)  
 INTEGRAL Likelihood = -4.25 Transmembrane 242 - 258 ( 241 - 259)  
 INTEGRAL Likelihood = -2.97 Transmembrane 54 - 70 ( 54 - 71)  
 INTEGRAL Likelihood = -2.87 Transmembrane 324 - 340 ( 319 - 340)  
 PERIPHERAL Likelihood = 0.16 133  
 45 modified ALOM score: 3.42  
 icml HYPID: 7 CFP: 0.684  
 \*\*\* Reasoning Step: 3  
 50 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6838(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15145 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 154/349 (44%), Positives = 220/349 (62%), Gaps = 6/349 (1%)

Query: 10 MSKKAQKIADVLEISVVLGIILGAILMLIFGYDPLWGYBGLFQTAPGSIKNIGEIFRAMPF 69

-1683-

M K+ + VPLI+++LG+ GR+IML+ GY GY L+ PG I +GE R+ P  
 Sbjct: 1 MVRKLSHLVPLIATILGLAAGALIMVSGYSVAGYSALWNGIPGEIYVGETTIQITP 60

Query: 70 LILIALGFSVASRAGFFNIGLPGQALSGMIANGWFALSHMPMPAMILCTIITIGIVAGG 129  
 IL L + A R G FNI+ GQ L GW AA W + D P + + I AGG  
 Sbjct: 61 YILSLGAVAFAPFTGLFNI+GVBGLLVGNIAVWVGTA+DGPAITHPLALITAAAGG 119

Query: 130 ITGAIPGILRAYLGTSEVITIMNNYIVLYGNAIVQVFFKSMIRTSIDSSVYVSANASY 189  
 + G IPGIL+A EVITIMNNYI L+ N I+ V D + + +AS  
 Sbjct: 120 LWSGIPGILKARFTVHEVITIMNNYIALAMINYIISNVLTIDH---QDKTKKHESASI 175

Query: 190 QTDWLSLNNRINIGIFIAIIAVLVWFLNKTITGFEIRSVGLNPNASEYAGMSAKR 249  
 ++ +L +T+ SR+++GI +A+A V++MF++NK+T GFE+R+VG N +AS+YACMS ++  
 Sbjct: 176 RSPFLGQITDYSRLHLGLIIVALLAIVMMFLINKSTKGFEIRAVGPNQASQYACMSVRK 235

Query: 250 TIILSMIISGAFAGLGGVVGELGTFFENVFVQPSLAIGFDGMAVSLAANSPIGILPAF 309  
 I+ SM+ISGAFAGL G+EGLTTE V+ + +GFDG+AV+LL N+ +G+ AA  
 Sbjct: 236 NIMISMIISGAFAGLAGAMEGLGTFFETAAVKAGFVGVDGIAVALLOGNTAVGVVLAAC 295

Query: 310 LFGVLSVGAPGSHI+AGIPPELIKVTASIIFFVGWHYIIEVYKPKKQ 357  
 L G L +GA N I +G+P E++ +V A II FV Y I +V+ K+  
 Sbjct: 296 LLGSLKIGALMFISSGVSPSEVDVIAIITILFVASSYAIRFVWGLKK 344

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2149> which encodes the amino acid sequence <SEQ ID 2150>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.74	Transmembrane	205 - 221 ( 200 - 228)
INTEGRAL	Likelihood = -12.42	Transmembrane	21 - 37 ( 14 - 45)
INTEGRAL	Likelihood = -7.22	Transmembrane	115 - 131 ( 111 - 135)
INTEGRAL	Likelihood = -4.78	Transmembrane	251 - 267 ( 249 - 269)
INTEGRAL	Likelihood = -2.50	Transmembrane	70 - 86 ( 69 - 86)
INTEGRAL	Likelihood = -2.34	Transmembrane	302 - 318 ( 300 - 318)
INTEGRAL	Likelihood = -1.44	Transmembrane	148 - 164 ( 147 - 165)
INTEGRAL	Likelihood = -1.33	Transmembrane	326 - 342 ( 326 - 342)

----- Final Results -----

bacterial membrane	---	Certainty=0.6095 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 293/358 (81%), Positives = 333/358 (92%), Gaps = 1/358 (0%)

Query: 6 RRRMSKNAQKIAVPLISVVLGAILGAIIMLIPGYDPLNGVGLFQTAFGSKNIGIEIFR 65  
 RR+ MSK AQKIAVPLISV+LG +LGAIM+IPGYD+NGVGLFQ AFGS+KNIGIEIFR  
 Sbjct: 6 RRRVMSKNAQKIAVPLISVLLGAILGAIIMVIGYDPIWGVGLFQIAFGSVKNIGIEIFR 65

Query: 66 AMGPILILIALGFSVASRAGFFNIGLPGQALSGMIAGNFWALSHMPMPAMILCTIITIGI 125  
 +MGPIILIALG+VASRAGFFN+GL QQAL+GMI+AGNFWAL+PMPMP +IL T +TG+  
 Sbjct: 66 SMGPIILIALGFTVASRAGFFNIGLPGQALAGMISAGNFWALNMPMPAMILCTIITIGI 125

Query: 126 VAGGITGAIPGILRAYLGTSEVITIMNNYIVLYGNAIVQVFFKSMIRTSIDSSVYVSA 185  
 +ACGI GAIPGILRAYLGTSEVITIMNNYI+LY GNAIVR +P+S+ ++ DS++ VS  
 Sbjct: 126 IACGIAGAIPGILRAYLGTSEVITIMNNYIIVYGNAIVQVGYPSVKGSDISTIQVSD 185

Query: 186 NASYQTWLSLNNRINIGIFIAIIAVLVWFLNKTITGFEIRSVGLNPNASEYAGM 245  
 NASYQT WLS+LNNRINIGIF AIIA+ L+WFLNKTITGFEIRSVGLNPNASEYAGM  
 Sbjct: 186 NASYQTHWLSLNNRINIGIFIAIIALVWFLNKTITGFEIRSVGLNPNASEYAGM 245

Query: 246 SAKRTIILSMIISGAFAGLGGVVGELGTFFENVFVQPSLAIGFDGMAVSLAANSPIGIL 305  
 S+KRTIILSMIISGA AGLGGVVGELGTFFENVFVQPSLAIGFDGMAVSLAANS+PI  
 Sbjct: 246 SSKRTIILSMIISGALAGLGGVVGELGTFFENVFVQPSLAIGFDGMAVSLAANSPIG 305

Query: 306 FAAFLEGVLSVGAPGSHI+AGIPPELIKVTASIIFFVGWHYIIEVYKPKKMGKK 362

-1684-

P++FLPGLV++GAPGNINAGIPPEL+KVVETASIIFFVG HY+IE Y+I+PKK +KGKG  
 Sbjct: 306 PSSFLPGLVNIAGAPGNINAGIPPELVVVETASIIFFVGSHYLIERTYIRPKKLVGKG 363

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1520

A DNA sequence (GBSx1610) was identified in *S. agalactiae* <SEQ ID 4675> which encodes the amino acid sequence <SEQ ID 4676>. This protein is predicted to be sugar ABC transporter, ATP-binding protein (mgfA). Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3851(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9803> which encodes amino acid sequence <SEQ ID 9804> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15144 GB:299120 similar to AEC transporter (ATP-binding protein) [Bacillus subtilis]  
 Identities = 311/497 (62%), Positives = 396/497 (79%), Gaps = 1/497 (0%)  
 Query: 14 VIEMKRITKFGDFVANDHINILVEKGEIHALLGENGAGKSTLMNMLAGLLEPTDSQIF 73  
 VIEH I K F VAND+INL V+KGEIHALLGENGAGKSTLMN+L GL P G+I +  
 Sbjct: 4 VIEMLNIRKAPPGIVANDNINLQVKVGEIHALLGENGAGKSTLMNVLPGLYQPERGIRV 63  
 Query: 74 NGQPYTIDSPSKSOLGIGMVHGHFPLVVEAFTVAENIVLGNETQNGVLDIKTAKEIKE 133  
 G+ V I+SP+K++ LGIGMVHGHFPLV+ PTVAENI+LG E + G+D K A +E+++  
 Sbjct: 64 RGEKVHINSFNKANDLGIGMVHGHFPLVDTTVAENIILGKEPKFGIRDRPAGQVQ 123  
 Query: 134 LSEKYGSLVNPNAKISDISVGAQQRVRLKTLRGADILIFDEPTAVLTTPSEIKELMTIM 193  
 +S+YGL +P AK +DISVQ QR BILKTLRGADILIFDEPTAVLT TP EIKEL IM  
 Sbjct: 124 ISDRYGLQIHPEAKADISVGMQQRABILKTLRGADILIFDEPTAVLTTPHEIKELMQIM 183  
 Query: 194 KSLVKGKSIILITHKLDEIRAVADKVTVIRKCKSIETVPVAGASSOOLAEHNVGRSVSF 253  
 K+LVKGKSIILITHKL EI +D+VTVIR+GK I+T+ V + +LA +MVGR VSF  
 Sbjct: 184 KNLVKGKSIILITHKLKEIMEICDRVTVIRKCKGIETVDRDVTQDELASLMVGRVSP 243  
 Query: 254 RTEKKEANPTDIILSVKDLVVEENRGGLVAVNKSLDVRAGRTVGIAGIDNGQGSRLQA 313  
 +TEK+ A P +L++ + V++ R G+ V++LGL V+AGRTVGIAG+DNGQGSRL+ A  
 Sbjct: 244 KTEKRAAQPGEVLAIDGITVKDIR-GIETVRLISLVKAGRTVGIAGIDNGQGSRLIPA 302  
 Query: 314 ITGLRKVTSGQIVIKKDKVTKPSRQITELSVGHVPEDRHRDGLVLDIMMAMNIALQY 373  
 +TGLRK SQ I + GK + + R+ITE +GH+P+DRH+ GLVLD + EN+ LQ+YY  
 Sbjct: 303 VTGLRKDTSGTITINGKQIGNLTPRKITSGIGHIPQDRHKGHLVLDFTIGENILLQSY 362  
 Query: 374 KEPLSHEKGIILNFAKIKETARQLMTEFDVRGAGSHVLARFSGGNOQKAI IAREVDRPDL 433  
 K+P S G+L+ ++ + AR L+TE+DVR E+ AR SGNQKAI RE+DR+PDL  
 Sbjct: 363 KKPYSALGVHLHGEMYKCARSLITTEYDVRTDREYTHARALSGNQKAIIGREIDRNPDL 422  
 Query: 434 LIVSQPTRGSLDVGAIIEYTHKRIEERDKGKAVLWVSPELDRIINLSDRIVAHVDRKQIGI 493  
 LI +QPTRGSLDVGAIIE+HK+LIE+RD GKAVL++SPEL+EI+NLSDRIAVI +G+I  
 Sbjct: 423 LIAAQPTRGSLDVGAIIEFVHKKLEIQRDAGKAVLLSPELREIMNLSDRIVAFEGRIAS 482  
 Query: 494 VKPDQTNKQELGIIMAG 510  
 V P +T +QELG+IMAG

-1685-

Sbjct: 483 VNFQETTBQELGLLMAG 499

Identities = 75/242 (30%), Positives = 128/242 (51%), Gaps = 24/242 (9%)

5 Query: 280 G/LAVKNLSLDVRAGEIVGLAGIDNGQSELIQATITGLRKVTSSQIVIKGKDVTKPSSRQ 339  
G++A N++L V+ GEI + G +G G+S L+ + GL + G+I ++G+V S +  
Sbjct: 16 GIVANENINLQVKKGKEITHALGENGAGKSTIMNVLPGLYQPEGEIRVRGEKVIHNSFNK 75

10 Query: 340 ITLESVGHVPEDEHRDGLVLD-MTMAENALQTYTYKEPLSHKGIINFPAKI--KEYARQLM 396  
+L +G V H+ +++D T+AEN+ L KEP F +I K +++  
Sbjct: 76 ANDLSIGMV---HQHFMVLDVTFVAENIILG---KEPKK-----FGRIDRKRAQSEVQ 122

15 Query: 397 TEPDVRGAGEHVLARG--FSGGNQKAIAREVDREDPDLILVSQPTLGL--DWGAIEYI 451  
D G H A+ S G QQ+A I + + R D+LI +PT L ++ + I  
Sbjct: 123 DISDRYGLQIHPEAKAADISVGMQQRARILKTLVRGADILIFDEPTAVLTPHEIKELMQI 182

20 Query: 452 HKRLIERDGGKAVLVVSELDLSILNLSDRIAVIHDKIQIVKGDQTNKQELGLIMAGG 511  
K L+E GK+++++ +L EI+ + DR+ VI GK + TN+ EL LM G  
Sbjct: 183 NMNLVKE---GKSIILITHKLKIMEICDRVTVIRRGEGIKTLVDRDYNQDELASLVGR 239

25 Query: 512 KI 513  
++  
Sbjct: 240 EV 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4677> which encodes the amino acid sequence <SEQ ID 4678>. Analysis of this protein sequence reveals the following:

Possible site: 60

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3558 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 431/511 (84%), Positives = 467/511 (91%), Gaps = 1/511 (0%)

Query: 10 MTQNVIMKEITTKPGDPVANDHINILTVKEGEIHALLGENGAGKSTLMMAGLLEPTG 69  
MTQ+VIM+RITTKPGDPVANDHINIL V KGEIHALLGENGAGKSTLMMAGLLEPTG 69  
Sbjct: 7 MTQNVIMKEITTKPGDPVANDHINILVNRKGEIHALLGENGAGKSTLMMAGLLEPTG 66

40 Query: 70 QIPINQCPVTIDSPSKSSQLGIGVHQHFMVLEAFTVAENILVGNRETTQGVLDITAK 129  
+I IN +PV IDSPSI9++LIGVHQHFMVLEAFTVAENI+LGNE +NG LD+ A+K  
Sbjct: 67 EIVINDKPVQIDSPSKSAKLIGVHQHFMVLEAFTVAENILVGNRETTQGVLDITAK 126

45 Query: 130 EIKELSEKYGSLVNPNAKISDISVGAQQRVEILKTLVRGADILIFDEPTAVLTPHEIKEL 189  
+IK LSEKYGSL+NP+AK+SDISVGAQQRVEILKTLVRGADILIFDEPTAVLTPHEIKEL 189  
Sbjct: 127 DIKVLSEKYGSLNPSAKVSDISVGAQQRVEILKTLVRGADILIFDEPTAVLTPHEIKEL 186

50 Query: 190 MTIMKSLVKBGKSIILITHKLDEIRAVADKVTVIRRGKSIETVPVAGSQQLAEH+VGR 249  
MTIMK+LVKKBGKSIILITHKLDEIRAVAD+VTVIRRGKSIETV VAGA+SQ LAEH+VGR 249  
Sbjct: 187 MTIMKSLVKBGKSIILITHKLDEIRAVADRVTVVIRRGKSIETVDVAGTSQQLAEH+VGR 246

55 Query: 250 SVSPFTEKKEANPTDILISVKDLVVEENRGGVAVKGLSLDVRAGEIVGLAGIDNGQSE 309  
SVSPF T KK A P D++LS+K+L +ENR GV AVK LSLDVRAGEIVGLAGIDNGQSE 309  
Sbjct: 247 SVSPFTSKKAEPIVDVILSKNLEVDENR-GVPAVKGSLDVRAGEIVGLAGIDNGQSE 305

60 Query: 310 LQIATGLRKVTSSQIVIKGKDVTKPSSRQITLESVGHVPEDEHRDGLVLDMTVAENAL 369  
LQIATGLRKV SG I+K +VT SSR+ITLESVGHVPEDEHRDGL+LD+++AEN AL 369  
Sbjct: 306 LQIATGLRKVSGSIMIKNEVTHLSRKITLESVGHVPEDEHRDGLDILISLAENAL 365

65 Query: 370 QTYTYKEPLSHKGIINFPAKIKEYARQLMTEPFDVRGAGEHVLARGPSGNGQKAIAREVDR 429  
QTYK+PLS GILN+ KI +YARQLM EFDVRA E V ARGPSGNGQKAIAREVDR 429  
Sbjct: 366 QTYTKPLSNGILNYTKINDYARQLMKEFDVRGAGNELVPARGPSGNGQKAIAREVDR 425

Query: 430 DPELLIVSQPTRGSLDVAIEYIHKRLIERDGGKAVLVVSELDLSILNLSDRIAVIHDKG 489

-1686-

DPDLLIVSQPTRIGLDVGAIRYHKRLI+ERDKGKAVIVVVSFELDEIINLSDRIAVIHGK  
 Sbjct: 426 DPDLLIVSQPTRIGLDVGAIRYHKRLI+ERDKGKAVIVVVSFELDEIINLSDRIAVIHGK 485

Query: 490 IGGIVKPOCTNKGELGILMAGGSIHKBERDV 520  
 IGGIV P+ TNKGELGILMAGS I KEE V  
 Sbjct: 486 IGGIVSPENTNKGELGILMAGGSIHKBEQHV 516

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 1521

A DNA sequence (GBSx1612) was identified in *S.agalactiae* <SEQ ID 4679> which encodes the amino acid sequence <SEQ ID 4680>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)  
 [Bacillus subtilis]  
 Identities = 164/335 (48%), Positives = 224/335 (65%), Gaps = 10/335 (2%)

Query: 18 LAACSHRGASKSGGKS-DSLKVAMVITDVGVDKSFNQSGWESMQAMGKQKGLKKA-GF 75  
 L ACQ+ S G+ VAMVITDVGVDKSFNQSGWESMQAMGKQKGLKKA G+  
 Sbjct: 11 LGACQNSGKSSGSGGKNSFVAMVITDVGVDKSFNQSGWESMQAMGKQKGLKKA 70

Query: 76 DYPQASBSDYATNLOTAVSSGYKLIPGIGFLHDAIDKADNNKDNVYIVVDVVIKGD 135  
 DY QS S++DY TNL+ + LI+G+G+ + D+I + AD K+ N+ I+D V+ KD  
 Sbjct: 71 DYLQSKSDADYTTNLNKLARENFDLYGVGLMRDSISEIADQKRNINFAIDAVVD-KD 129

Query: 136 NVASVVPADNESAYLAGIAAAKTKTKTVGVFGVGMNESEVITREFEGFAGVSKVDSIKI 195  
 NVAS+ F + E ++L G+AAA ++K+ +GFVGMNESE+I +FE GF AGV++V+ +  
 Sbjct: 130 NVASITPKQEGSFLVGVAALSKSGKGI GFVGMNESELIKFEVGFAGVQAVNPKAVV 189

Query: 196 KVDYAGSFGDAAGKGTIAAAQYASGADIVYQAGGTGAGVPSRAKSNESLKEALDKWVL 255  
 +V YAG F A GK R + Y SG D++Y AG TG GVF+BAK+ + + D VWV+  
 Sbjct: 190 EVKYAGGFDKADVGKATAGNYKSGVDVYHSAGATGTGVTBAIHLKCKEDPKRD-VWV 248

Query: 256 GVDKQABGKYTKSDKASNPVLAASSIKVSGVELIATKTSKGFPGGNVITYGLKD 315  
 GVD+DQ ABG+ +G N L S +K+V VE + K S GKFFGG TRL  
 Sbjct: 249 GVDKQYABGQV---EGTDNVTLTSMWKKVDTVVEDVTKKASDGKFPQGGITLYGLQD 305

Query: 316 GVDIATT--NLSDDAVKAIKEAKLIISGDIKVP 348  
 GV I+ + NLSDD +K+ + K KII G +++P+  
 Sbjct: 306 GVGISPSKQLNSDDVIKAVDKWKKLIIDG-LKIPA 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 861> which encodes the amino acid sequence <SEQ ID 862>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.



-1687-

Identities = 275/351 (78%), Positives = 312/351 (88%), Gaps = 3/351 (0%)

Query: 1 MNKKIAGIGLASIAVLSAACGHRGASKSG--GKSDSLKVAMVTDGVDKSPNQSGWE 58  
 5 Sbjct: 1 MNKK G+GLAS+AVLSIAACG+RGASK G GK+D LKVAMVTDGVDKSPNQSG WE  
 MNKKFIGIGLASIAVLSIAACGHRGASKSGGASGTD-LKVAMVTDGVDKSPNQSGWE 59

Query: 59 GMQANGKKNLKKGAGFDYFQSGASESDYATNLDTAVSSGYKLIFGIGFSLHDAIDKAADN 118  
 G+Q+WGK+ GL+KG GFDYFQSG SES+YATNLDTAVS GY+LI+GIGF+L DAI KAA +  
 10 Sbjct: 60 GLQSNGBWMLQKGTGFDYFQSGTSESEYATNLDTAVSGGYQLTYGIGFALKDAIDKAAGD 119

Query: 119 NKDNYVIVDDVIKGDNVASVVPADNESAYLAGIAAAKTTKTKTVGPGVGMSEVITRF 178  
 N+ V+VI+DD+I+G+DNVASY PAD+E+AYLAGIAAAKTTKTKTVGPGVGMSE VITRF  
 15 Sbjct: 120 NEGVKFLIIDDIIKGDNVASVTFADHEAAYLAGIAAAKTTKTKTVGPGVGMSEVITRF 179

Query: 179 EKGFEAGVKSVDSIKIKVDYAGSPGDAAGKGTIAAAQYASGADIVYQVAGGTGAGVPE 238  
 EKGFEAGVKSVSD +I++KVDYAGSPGDAAGKGTIAAAQYA+GAD++YQ AGGTGAGVPAE  
 20 Sbjct: 180 EKGFEAGVKSVDDTIQKVDYAGSPGDAAGKGTIAAAQYAGADIVYQVAGGTGAGVPE 239

Query: 239 AKERNESLKEADKVVVLGVDRDQAAEGKYTSKDGKASNFVLASSIKEVGKVELIATKTS 298  
 AK+ N3 EADKVVW+GVDRDQ EGKYYTSKDGK +NVLASSIKEVGK+V+LI ++  
 25 Sbjct: 240 AKAINERSEADKVVVIGVDRDQDEGKYTSKDGKASNFVLASSIKEVGKAVQLINKQA 299

Query: 299 KGKFPGRNVITYGLKDGSDVIATNLSDDAVKAKEAKAKISGDIKVPK 349  
 KFPGR T YGLKDGSD+IATNLW+S+AVKAKEAKAKI SGGINKVP K  
 30 Sbjct: 300 DKKFPGRKITYGLKDGSDVIATNLW/SKZAVKAKEAKAKISGDIKVPK 350

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 9061> which encodes amino acid sequence <SEQ ID 9062>. Analysis of this protein sequence reveals the following:

Possible site: 17

30 >>> May be a lipoprotein

----- Final Results -----

35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 414 bits (1052), Expect = e-117  
 40 Identities = 196/347 (56%), Positives = 253/347 (72%), Gaps = 2/347 (0%)

Query: 1 MNKKVMSLGLVSTALFTLGCGTNNSAKT--TNSLKITAMITNQTGIDDKSPNQSGWEG 58  
 Sbjct: 1 MNKK+ +GL S A+ +L C + A ++ +SLK+AM+T+ G+DDKSPNQSG WEG+ 60

45 Query: 59 QAWGKNLKLKGGYDYFQSANSEPTTNLESATVNTGYNLVPFQIGFPLHDAVEKVAANNP 118  
 QAWGK+N L+KG G+DYFQSA+ES++ TNL++AV++GY L+FGIGF LHDA++K A NN  
 Sbjct: 61 QAWGKNLKLKGAAGFDYFQSANSESDYATNLDTAVSSGYKLIFGIGFSLHDAIDKAADN 120

50 Query: 119 DNHFAIVDDVIKGRNVASITPSDEHAAYLAGVXXXXXXXQVGFVGMSEGVVVRFEK 178  
 D ++ IVDENVKGR+ NVAS+ F+D+B+AYLAG+ VGFVGMSE +V+ RFRK  
 Sbjct: 121 DNVYIVDDVIKGRNVASVVPADNESAYLAGIAAAKTTKTKTVGPGVGMSEVITRF 180

55 Query: 179 GFEAGVKSVDDTIKRVAYAGSFXXXXXXXXXXXXKRGADIVYHAAGGTGAGVPSAK 238  
 GFEAGVRSVD +IK++V YAGSF QAD++Y AGGTGAGVPSAK  
 Sbjct: 181 GFEAGVKSVDSIKIKVDYAGSPGDAAGKGTIAAAQYASGADIVYQVAGGTGAGVPSAK 240

60 Query: 239 SINEKREEDKVVVIGVDRDQSDGKYYTKDGKSNFVLTSIKEVGKALVKVAKTSED 298  
 S NE KE DKVW+GVDRDQ+ +GKTY+KDKK++NVL SSIKEVGK++ +A KTS+  
 Sbjct: 241 SRNESLKEADKVVVLGVDRDQAAEGKYTSKDGKASNFVLASSIKEVGKVELIATKTSK 300

Query: 299 QFPGQIITTFGLKGGVSLTDTALITQDTXXXXXXXKGTITFP 345  
 +FPGG +TT+GL+GGV + T L+ D G I VP  
 65 Sbjct: 301 KFPGRNVITYGLKDGSDVIATNLSDDAVKAKEAKAKISGDIKVP 347

-1688-

SEQ ID 4680 (GBS211) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 6; MW 40kDa).

The GBS211-His fusion product was purified (Figure 205, lane 8) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 259A) and FACS (Figure 259B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1522

A DNA sequence (GBSx1613) was identified in *S.agalactiae* <SEQ ID 4681> which encodes the amino acid sequence <SEQ ID 4682>. This protein is predicted to be cytidine deaminase (cdd). Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2112(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9801> which encodes amino acid sequence <SEQ ID 9802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB51906 GB:AJ237978 cytidine deaminase [Bacillus psychrophilus]  
Identities = 66/114 (57%), Positives = 81/114 (70%)

Query: 26 KASENAYVPYSKFPVGAALKTAEKIPTCNVENISYGLANCAERTAIKFAVSEGYKDFS 85  
KA E AYVPYSKFPVGAAL +G I+ GGN+EN +Y + NCAERTA FKAUS+G + F  
Sbjct: 12 KAREQAYVPYSKFPVGAALLAEDGTIYEGCNIENSAYSMNCAERTAFKAVSDGVRSPK 71

Query: 86 EIAIVGNTERPISPOGACRQVMVEFFPNQAKVTLIARNGKIVETTVGELLPSYF 139  
+A+ +TE P+SPGACROV+ EF N + V L G ETTV +LLP +F  
Sbjct: 72 ALAVVADTEGFPVSPGACRQVIAEFCNGSMFVYLNLGQDIETTVAKLLPGAF 125

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4683> which encodes the amino acid sequence <SEQ ID 4684>. Analysis of this protein sequence reveals the following:

Possible site: 13  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0041(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15143 GB:Z39120 similar to ABC transporter (lipoprotein)  
[Bacillus subtilis]  
Identities = 152/339 (44%), Positives = 223/339 (64%), Gaps = 11/339 (3%)

Query: 8 IGLVSTALPFTLGECITNN---SAKQTITNLSLKIAMITNQTGIDDKSPNQSAWEGLQAWGKE 64  
+LV A LG C N+ S N +A+T+ G+DDKSPNQSAWEG+QA+GKE  
Sbjct: 1 MSLVIAAGTILGACNSEKSGSGSGEYKPKFSVAMVTDVGGVDKSPNQSAWEIGQAFQRE 60

-1689-

Query: 65 NKLEKRGK-GYDYPQSANESSFTTNLESAVINGYNLVFGIGFPLHDAVEKVAANNPDNHPA 123  
 N L+RGK GYDY QS +++++TINL ++L+G+G+ +D++ +A+ +FA  
 Sbjct: 61 NGLKKGKNGYDYLQSKSDADYTTINLKLARENFDLIYGVGLMREDSISIAIDQRKNTNFA 120

5 Query: 124 IVDVVIRKQKNVASITFSDEHAYLAGVAAAKTTKTQKQVFGVGMGGVVKRFKFGFSAQ 183  
 I+D V+ + NVASITF + E ++L GVAAA ++K+ ++GFGVGMG +++K+FE GF AG  
 Sbjct: 121 IIDAIVD-KDNVASITFKQKQSFVLGVAAALSSKSGKLGFGVGMSELIKFKFEVGFAG 179

10 Query: 184 VKSVDDTIKRVAYAGSPADAANGKTIAAAQYABGADVITHAAGTGGMVFSEAKSINEK 243  
 V++V+ V V YAG F A GK A + Y G DVITH+AG TG GVF+BAK++ ++  
 Sbjct: 180 VQAVNFKAVVEVKTAGGDFKADVGKATABSMYKSGVDVITHSAGATGTGVTFAKNLAKG 239

Query: 244 RKESDKVWVIGVDRDQSEDKYTTKDGKSANFVLTSISIEVGKALVKVAKTSBDQFFGG 303  
 + D VVWIGVD+DQ +G+ +G N LTS +K+V + V K S+ +PFGS  
 Sbjct: 240 DPARD-VWVIGVDKQYABGV---EGTDDNVLTSMVKVDTVVEDVTKASDGKFFGG 295

15 Query: 304 QITTFGLKEGGVSLITDA--LTQDTKKAIEAAKKAIIIG 340  
 + T+G+ +GV ++ L+ D KA++ KK II+G  
 Sbjct: 296 ETLTYGLDDQGVSGSPKQNLSDVKAVDKWKKKIIG 334

20

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/128 (68%), Positives = 107/128 (82%)

25 Query: 15 MGNIELAKLVAKASRDYVFPYSKFPVGAALRTABGKIFTGNCVENISYGLANCAERTAF 74  
 MG +L AV+ASE AYVYS FVGAAL+T +G I+TGCN+EN+S+GL NC ERTAF  
 Sbjct: 1 MGTDLVSCAVQASEYATVPYSHFPVGAALRTKDTITTCNINENVSGLTNGERTAF 60

Query: 75 KAVSBGYKDFSEIALYGNTERFISPOGACRQVMVEFFNNQAKVTLIAKNGKTVETTVGEL 134  
 K+H+S+G+K+ EIALY T +P+POGACRQVM EFF+ ++ VTILAKNG+TVE TVG+L  
 30 Sbjct: 61 KAISDGKELVIEIALYGETMQPVSPGACRQVMVEFFDPGSLVTILAKNGQTVETTVGEL 120

Query: 135 LPFSFVDL 142  
 L YSF DL  
 35 Sbjct: 121 LLFSFVDL 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1523

40 A DNA sequence (GBSx1614) was identified in *S. agalactiae* <SEQ ID 4685> which encodes the amino acid sequence <SEQ ID 4686>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2979 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9799> which encodes amino acid sequence <SEQ ID 9800> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11882 GB:Z99104 alternate gene name: ybaA-similar to  
 hypothetical proteins [Bacillus subtilis]  
 Identities = 90/201 (44%), Positives = 144/201 (70%), Gaps = 5/201 (2%)  
 55 Query: 1 MANVYVTENPNVEHDIHLAVKLLCGSPSELTAGVPSKRMIDYGGQVLLNSLHP-EENK 59  
 M+ Y+H+ P+V+ + + +L + P+P +D+GVFSK+ +D+CS+L+H+S E  
 Sbjct: 1 MSBHYYSSEKPEVKNKQVTSFRLNKDPTPTSDSGVPSKGBVDVFCGRLLIDSFEZFEVQ 60

-1690-

Query: 60 SLDLGGCGYGLGISLAK-VQGVKATMVDINTRALELAKKQATNGVV-VVFQSNYIEN 117  
 +LD+GGGYGP+G+SLA + M+D+N RA+EL+ +NA +NG+ V+++QS+++ N  
 Sbjct: 61 GILDVGGYGPGLGISLASDFKDRTHMIDVNERAVELSNERRASQNGITNVKIYQSLFNS 120

5 Query: 118 I--SRTFDYIISNPPIRAGKVVHSIIESICYLMTGSSLTIVIQKQGAPSAKAKMLDT 175  
 + ++TF I++NPPIRAGK+VVR+I E+S +L G L IVIQKQGAPSA K+ +  
 Sbjct: 121 VDSAGTFAISILTNPPPIRAGKVVHAIPEKSAEHLKASGELMTIVIQKQGAPSAIEKLEEL 180

10 Query: 176 FGNCIDILEKDKGYIILRSEKV 196  
 F +++K KGYI++++KV  
 Sbjct: 181 FDESVVQKKKGYIILRAKV 201

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4687> which encodes the amino acid sequence <SEQ ID 4688>. Analysis of this protein sequence reveals the following:

15 Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4232 (Affirmative) < succ>  
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 139/195 (71%), Positives = 165/195 (84%)

25 Query: 1 MANNYYTENPVEHIDHILNVLLGESFSLTDAGVFSKRMIDYGSQVLLNSLHFEKKKS 60  
 M MYF ENP+ HDIHE+ V+LL F+FLTD+GVFSK+N+D+GSQVLL +L+F++N+  
 Sbjct: 12 MTFMYIDENPDSLIDHIEVKVLLNHPPTFLTDGVSFSGVDFGSQVLLKTLNFKENR 71

30 Query: 61 LLDLGGCGYGLGISLAKVQGVKATMVDINTRALELAKKQATNGVVVFQSNYIENISK 120  
 +LDLGGCGYGLGISLAKVQ V AT+VDIN RAL+LA+KNAI N V V +FQSNYIENIS  
 Sbjct: 72 VLDLGGCGYGLGISLAKVQGVKATMVDINRALDLARQATNMQVATTFQSNYIENISG 131

35 Query: 121 TFDYIISNPPIRAGKVVHSIIESICYLMTGSSLTIVIQKQGAPSAKAKMLDTFGNCD 180  
 F++IISNPPIRAGK+VVR+IE+ST +L G LTVIQKQGAPSAKAKM KGN +  
 Sbjct: 132 HFEHIIISNPPIRAGKRVVHSIIEKSIDPLVWNGDLTVIQKQGAPSAKAKMATIFGNVE 191

Query: 181 ILKDKKGYIILRSEK 195  
 IL+KDKGYI+LRS K  
 40 Sbjct: 192 ILRDKKGYIVLRSEK 206

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1524

45 A DNA sequence (GBSx1615) was identified in *S.agalactiae* <SEQ ID 4689> which encodes the amino acid sequence <SEQ ID 4690>. This protein is predicted to be pantothenate kinase (coaA). Analysis of this protein sequence reveals the following:

possible site: 28  
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5021 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06594 GB:AP001516 pantothenate kinase [Bacillus halodurans]  
 Identities = 140/307 (45%), Positives = 203/307 (65%), Gaps = 5/307 (1%)

-1691-

Query: 4 EFINFDRISRENWKDLHQSQALLTEKLESISKLNNDNINIQDVIDIYPLINLIQIYKR 63  
 +F + +SR WK L + S + E+KLE + LN+ I + +V DIY+PL L+ ++  
 5 Sbjet: 8 DFFFTYTLVSRGQWKSFLKASSLPINBQLEQLVGLNRPITLNEKVDIYVPLIALRLHVAT 67

Query: 64 SQENLFSFKAIPLKKNRYQRPFIIGISGSVAVGKSTTSRLQLLISRTFKDSHVELVTTD 123  
 + + L K F + FTIG+GSVAVGKST+RLQ L+ + HV+LVTTD  
 Sbjet: 68 AYQLRQQQKRGPFHIGKNSRPFITIGAGSVAWGKSTIARLLQKLKAWPEHHRVDELVTTD 127

10 Query: 124 GFLYPNFKLIQNGILNRKGPFESYDMSLLNFDLTIKNGIT-AKIPIYSHRIYDIVPNQL 182  
 GFLYPN L G++++KGPFESYD+ +L+ FL +K G K P+YSH Y+IV  
 Sbjet: 128 GFLYPNLTLEAGLMDKKGPFESYDLPALIRFLSDVKNRGPYKAPVYSHLTYNIVEGDY 187

15 Query: 183 QTIETPDFLILEGINVFO-QQNHRL---YNNDYDFDSIYDAENKQIBEWYIQRPNLSL 238  
 Q + PD +I+EGIN Q N++NH + +++D+PDFSIY+DA+ +QI +WY+RF L  
 Sbjet: 188 QVVHEPDYIVYBSINVLQVKNRHHIPNVFVSDFDFSIYVDAKESQLWYIERFKLIQ 247

20 Query: 239 QLABADPSNYHKTQIPPHKAMELAKDIWKTINLVNLEKYIEPTNRADPIIHKGHHK 298  
 A DP++Y+H+F + +A +A IWK IN VNL+ I PT++RAD ++ KG HH  
 Sbjet: 248 NTAQDPNSYPIRPHFLSEVEAQFATSIWKNINLVNLEHNLPTKRAIDLQKGPFRH 307

Query: 299 IDETYLK 305  
 IDE+ L+  
 Sbjet: 308 IDEVKLR 314

25

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4691> which encodes the amino acid sequence <SEQ ID 4692>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4790 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 219/306 (71%), Positives = 269/306 (87%)

Query: 1 MNNEFINFDRISRENWKDLHQSQALLTEKLESISKLNNDNINIQDVIDIYPLINLIQI 60  
 M+NEFINF++ISREWK LHQ+++ALLT++EL+SI SLNDNI+I DVIDIYPLINLIQ+  
 40 Sbjet: 1 MSNEFINFEKISRESWKTLHQKAKALLTCEELKSTISLNDNISINDVIDIYPLINLIQV 60

Query: 61 YKRSQENLFSFKAIPLKKNRYQRPFIIGISGSVAVGKSTTSRLQLLISRTFKDSHVELV 120  
 YK +QENLFSFK++FLKK+ RPTIIGISGSVAVGKSTTSRLQLL+SR +S VVELV  
 45 Sbjet: 61 YKIAQENLFSFKSLFLKKDIQLRPFITIGISGSVAVGKSTTSRLQLLERTHPSNQVEIV 120

Query: 121 TIDGFLYPNFKLIQNGILNRKGPFESYDMSLLNFDLTIKNGITAKIPIYSHRIYDIVPN 180  
 TIDGFLYPN+ LI+ G+LNKRGPFESY+ME LL+FLD+IKNG TA P+YSH+IYDI+PN  
 Sbjet: 121 TIDGFLYPNQLPILQGLILNRKGPFESYDMSLLNFDLDSIKNGQTAFAPVYSHDIYDIPI 180

50 Query: 181 QLQTIETPDFLILEGINVFOQNHRLYNNDYDFDSIYDAENKQIBEWYIQRPNLSL 240  
 Q Q+ PDFLI+GGINVFOQNHR+Y+M+DYDFDSIYDA++ IR WY+RF S+L+L  
 Sbjet: 181 QKQSFNNPDFLIVBSINVFQNNRNLMSDYDFDSIYDAENKQIBEWYIQRPNLSL 240

55 Query: 241 ASADPSNYHKTQIPPHKAMELAKDIWKTINLVNLEKYIEPTNRADPIIHKGHHK 300  
 A+ DP NYY + Q+P +A+ A+++WKT+NL NLEK+IEPTNR+ I+HK HKID  
 Sbjet: 241 AKRDHNYAQAQLRSEALAPARNVWKTINLVNLEKPIETPNRRLILHKGADHKID 300

Query: 301 EIYLEK 306  
 EIYLEK  
 60 Sbjet: 301 EIYLEK 306

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1692-

**Example 1525**

A DNA sequence (GBSx1616) was identified in *S. agalactiae* <SEQ ID 4693> which encodes the amino acid sequence <SEQ ID 4694>. Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3866 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05058 GB:AP001511 ribosomal protein S20 (BS20) [Bacillus halodurans]  
 Identities = 47/86 (54%), Positives = 59/86 (67%), Gaps = 7/86 (8%)  
 Query: 3 VKTLANIKSAIKRAELAVKQNEKNSAQKSMRTAIKAFEA---NPSEELYRA---ASSS 55  
           +K ANIKSAIKR + N K+ +N+ KSA+RTAIK FEA N E +A A+  
 Sbjct: 1 MKGNANIKSAIKRVTNEKRRRIQNASVKSALRTAIKQFEAKVENNDAAKAAAFVEATKK 60  
 Query: 56 IDKRAASKGLIHNTKASRDKARLATKL 81  
           +DKAA+KGLIH N ASR K+RLA KL  
 Sbjct: 61 LDKAANKGLIHGNAASRQKSLAKKL 86

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4695> which encodes the amino acid sequence <SEQ ID 4696>. Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3872 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 76/82 (92%), Positives = 78/82 (94%)  
 Query: 1 MEVKTLANIKSAIKRAELAVKQNEKNSAQKSMRTAIKAFEAANDPSEELYRAASSSIDYAA 60  
           +EVKTLANIKSAIKRAELAVK NEKNSAQKSMRTAIKAFEAANDPSEELYRAASSSIDYAA  
 Sbjct: 1 LEVKTLANIKSAIKRAELAVKANEKNSAQKSMRTAIKAFEAANDPSEELYRAASSSIDYAA 60  
 Query: 61 SKGLIHNTKASRDKARLATKL 82  
           SKGLIH NKASRDKARLA KL  
 Sbjct: 61 SKGLIHNTKASRDKARLATKL 82

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1526**

A DNA sequence (GBSx1617) was identified in *S. agalactiae* <SEQ ID 4697> which encodes the amino acid sequence <SEQ ID 4698>. Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence  
           INTEGRAL Likelihood = -10.99 Transmembrane 31 - 47 ( 25 - 51)  
 ----- Final Results -----  
           bacterial membrane --- Certainty=0.5394 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1693-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:AAC35851 GB:AF086736 amino acid-binding protein Abp

[Streptococcus uberis]

Identities = 169/269 (62%), Positives = 203/269 (74%), Gaps = 2/269 (0%)

Query: 29 KNILLTIIIFOLFMIILSACMSNKRMAQIDNWEHYQKKKKITIGFDNTFVPMGFERSGD 88

K ILLT + + L ACG S+ A D W+ Y+KEK IT+GFDNTFVPMGF+ SG

Sbjct: 4 KKILLTTTALASTLFLVACGKSSA--AKTDQMDTYKKKKSITLGFNTFVPMGFKDESGK 61

Query: 89 YTGFDIDLANAVFKEYGISVKWQPINWDMKETEANNQINIDLIWNGYSKTERAKKVAFTN 148

TGFD++LA AVF+SYGI VK+QPINWD+KETEL NG ID+IWNQYS T ER KVAF+

Sbjct: 62 NTGFDVELAKAVPQYGIKVKFQPINWDLKETELNGKIDMIWNGYSVTKERQAKVAPST 121

Query: 149 PYNRRHQVIVTKTSSHINSIKMKHGKLGAGSSSGFDAPNAKPDILKKFVGEKAVQYD 208

PYN N QV+VTK SS+I S MKKG LGAGSSSG+DAF + P +LK VK +A QY+

Sbjct: 122 PYNKNEQVLVTKSSNITSFAMKKGVLGAGSSSGYDAFTSNPKVLKIDIVKNDATQYE 181

Query: 209 TPTQALIDKNNRIDGLLIDDEVYANYTLKQEGNIKAYYPVKTAYQGENFVVGARKVDRL 268

TF QA IDLKN+RIDGLLID+VYANYTLKQEG + Y VK+ +GE+F VG RK D+ L

Sbjct: 182 TPIQAFIDLNDRIDGLLIDKVVANYTLKQEGELIANYINIVKSEFDGEDFAVGVRKEDKIL 241

Query: 269 IEKINKAFQLNEKGRFQKISYKWFGEVD 297

++ IN AF +L+ G+FQ+IS KWFGEVD

Sbjct: 242 LKNINSAPTKLYTKGKFQGISQKWFGEVD 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4699> which encodes the amino acid sequence <SEQ ID 4700>. Analysis of this protein sequence reveals the following:

Possible site: 21

&gt;&gt;&gt; May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the databases:

&gt;GP:AAC35851 GB:AF086736 amino acid-binding protein Abp

[Streptococcus uberis]

Identities = 176/277 (63%), Positives = 220/277 (78%), Gaps = 1/277 (0%)

Query: 1 MIIKRTVAILAIASSFFLVAQATKSLKSGDAWGVQKQKSITVGFNTFVPMGFKDES 60

M +KK + LA+AS+ FLVAC + +K+ D W Y+K+KSIT+GFNTFVPMG+KDES

Sbjct: 1 MNLKILLITLALASTLFLVACGKSSAAT-DQMDTYKKKKSITLGFNTFVPMGFKDES 59

Query: 61 GRCKGFDIDLAKEVFHQYGLKVNFPQAINWDMKEARLNGKIDVINNGYSITKERQKVAF 120

G+ GFD++LAK VP +YG+KY FQ INWD+KE EL NGKID+IWNQYS+TKERQ KVAF

Sbjct: 60 GRNTGFDVELAKAVPQYGIKVKFQPINWDLKETELNGKIDMIWNGYSVTKERQAKVAF 119

Query: 121 TDSYMRNEQLIVVKRSDRTISDMKHKLGAQSSSGYDLSLLTPKLLKDFINQDANQ 180

+ YN+NEQ++V KK S+I + + MK KVLGAGS SSGYD+ PE+LKD +K+ DA Q

Sbjct: 120 STPYMRNEQVLVTKSSNITSFAMKKGVLGAGSSSGYDAFTSNPKVLKIDIVKNDATQY 179

Query: 181 YETFTQAIFDLKSDRIDGLIDKVVANYTLKQEGELIANYINIVKSEFDGEDFAVGVRKEDK 240

YETF QAFIDLK+DRIDG+LIDKVVANYTL +EG+L NY ++ + P+ E F+VG+RKEDK

Sbjct: 180 YETFIQAFIDLNDRIDGLIDKVVANYTLKQEGELIANYINIVKSEFDGEDFAVGVRKEDK 239

Query: 241 TLQAKINRAFRVLYQNGKQFQAISEKWFQGDVATANIK 277

L IN AF LY+ GKQF IS+KWFQ+DVAT N+K

Sbjct: 240 TLKKNINSAPTKLYTKGKFQGISQKWFQGDVATNRVK 276

-1694-

An alignment of the GAS and GBS proteins is shown below.

Identities = 151/266 (56%), Positives = 189/266 (70%), Gaps = 4/266 (1%)

```

Query: 32  LLTIIFGLFMILSACGMSNKEMAGIINWHEYQKEKKITIGFDNTFVPMGFESRSQDITG 91
      5  +L I   F++   AC + K +   D W YQK+K IT+GFDNTFVPMG++ SG G
      10  ILAIASSFFLV---AC-QATKSLKSGDANGVIYQKQKSTVGFDMTFVPMGYKQBSGRCK3 65

Query: 92  FDIDLANAUFVEKYGISVVGQPINWDMKSTELANGHIDLIWNGYSKTAERAKKVAFTNPVM 151
      10  FDIDLA VF +YG+ V +Q INNDMKE ELNNG ID+IWNNGYS T ER KVAFT+ YH
      66  FDIDLAKEVHFHYGLKVNFGQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYM 125

Query: 152  NNHQVIVTKTSSHINSIKMNGKKLGAQSGSSQFAPNAKPDILKKFVKGKEAVQVDITFT 211
      15  N Q+IV K S I +I DMK K LGAQS SSG+D+ P +LK F+K K+A QX+TFT
      126  RNHQIIVVKKRSIDI KTI SDMKHKVLGAQSSAGGYDLSLKTPEKLLKDFIKNKDANQYETFT 185

Query: 212  QALIDLNNRIDGLLIDEVYANYFLKQGNKAYFYVKTAYQGNFVVGARKVDRRLIEK 271
      20  QA IDLK++RIDG+LID+VIANYFL +BG ++ Y + T ++ E F VG RK D+ L K
      186  QAFIDLKSDRIDGILIDKVIANYFLAKEGQLENTYMIPTTFVMEAFVSVGLRKEDKTLOAK 245

Query: 272  INKAFKQLHNKGRFQKISYKWFQEDV 297
      25  IR+AF+ L+ G+FQ IS KWFQ+DV
      246  INRAFVLVYQNGKFPQALISERWFGDDV 271
  
```

A related GBS gene <SEQ ID 8833> and protein <SEQ ID 8834> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: 22  Crend: 4
      Sequence Pattern: CGMS
      SRFGLG: 0
      McG: Length of UR: 22
      Peak Value of UR: 3.05
      Net Charge of CR: 2
      McG: Discrim Score: 11.16
      GVH: Signal Score (-7.5): -1.96
      Possible site: 24
      >>> May be a lipoprotein
      Amino Acid Composition: calculated from 23
      ALOM program count: 0 value: 8.96 threshold: 0.0
      PERIPHERAL Likelihood = 8.96 68
      modified ALOM score: -2.29

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
  
```

The protein has homology with the following sequences in the databases:

```

62.2/75.8% over 270aa
GP|3603430| amino acid-binding protein Abp Insert characterized
Streptococcus uberis
ORF0904 (385 - 1203 of 1503)
GP|3603430|gb|AAC35851.1||AF086736(4 - 274 of 277) amino acid-binding protein Abp
[Streptococcus uberis]
%Match = 34.8
%identity = 62.1 %Similarity = 75.7
Matches = 169 Mismatches = 65 Conservative Sub.s = 37

153 183 213 243 273 303 333 363
FRYLGKGNVSH*LVK**LIHRLVMSQALALLIQSCVKK*KN*FYKIKQV*HKL**HMIPLNLKVYILRPSKILRL
393 423 453 483 513 543 573 603
GGRLLTHNNILLTIIFGLFMILSACGMSNKEMAGIINWHEYQKEKKITIGFDNTFVPMGFESRSQDITGFDIDLANAUF
  
```



-1695-

```

      : | ||| : : | || : | | : ||| ||:|||||||: | | |||:| ||
      MNLKLLITLALASTLFLVACGSS--AAKIDQNDITYKKKSTLGLFDNTFPVWGFKDESGKNTGPDVELAKAVF
      10      20      30      40      50      60      70

5      633      663      693      723      753      783      813      843
      KEYGISVWQFQINWDMKETELANGNIDLWNGYSKTAERAKKVAFTNPMNNHQVIVTKTSSHINSIKMGGKKLGAQSG
      :||| :|:|||||:||||| :| :|:||||| :| :|:||||| :| :|:||||| :| :|:||||| :| :|:|||||
      QYGIKVKFQPINWDLKETLKGKIDIMWNGYSVTKERQAKVAFSTPYMKNEQVLVTKKSSNITSFAMKNGVLGAQSG

10      90      100      110      120      130      140      150

      873      903      933      963      993      1023      1053      1083
      SSGPDAPNAKQPIILKKFVKGKRAVQYDTFTQALIDLKNNRIDGLLIDEVYANYLLKQSGNIKAYFPKTAQCGENFVVG
      :||| :|:||||| :| :|:||||| :| :|:||||| :| :|:||||| :| :|:||||| :| :|:|||||
      SSGYDAFTSNPKVLKIDVKLNDATQYETFTQALIDLKNNRIDGLLIDEVYANYLLKQSGELITNYIVKSEFGEQVAVGV

15      170      180      190      200      210      220      230

      1113      1143      1173      1203      1233      1263      1293      1323
      RKVDRLEIKINKAFKQQLNKGKRFQKISYKWFGEDEVYSKSKTYRNFS*SFILRKN*IKNIDISIVF*VN*PSLVARRALS
      || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
      RKECKILLKNNISAPTKLYKTKGFQESKWFGEDEVATENVKK

20      250      260      270

```

SEQ ID 8834 (GBS225) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 10; MW 32kDa). The GBS225-His fusion product was purified (Figure 205, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 266), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1527

A DNA sequence (GBSx1618) was identified in *S. agalactiae* <SEQ ID 4701> which encodes the amino acid sequence <SEQ ID 4702>. This protein is predicted to be arginine ABC transporter, ATP-binding protein (gInQ). Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3229(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AB49429 GB:U73111 glutamine transport ATP-binding protein GINQ
[Salmonella typhimurium]
Identities = 94/210 (44%), Positives = 146/210 (68%), Gaps = 3/210 (1%)

Query: 1 MLELKNISKCYQKEIFKDFNLTVEGKILSLVGSQGGKTTLLRMAGLEKIDSGTIVH 60
M+E KN+SK +G ++ +L +G+++ +GSGG QK+TLR+ LE+I SG ++
Sbjct: 1 MIEPKNVSKHFGPTQVLHNDLNRQSEVVVITGPGSSGSKSTLRLCKINLEITSGDLIV 60

Query: 61 DGKEVS---VDHLETLNLLGFPVQDFQAFPHLTVLQNLILSPVTMGLSKELAKEKALVL 117
DG +V+ VD G VQ F LPHLIT L+N++ P++ G+ KE A++A+ L
Sbjct: 61 DGLKVNDPKVDLERLIRQEGMVFQGFYLFPHLTALEVMFGPLRVGKKEEAKQAL 120

Query: 118 LERLGLADEHALVYPPSLGGQKQKVALARANNIDPOLIGYDEPTSLDPELRQELKIL 177
L ++GL +A YP LSGGQ+QKVA+ARA++ P+++ +DEPTSLDPELR IV K++
Sbjct: 121 LAKVGLAERAHYPSLGGQQKQKVAIARALAVKQMLFDEPTSLDPELRHVLKQV 180

Query: 178 QNRSTGMTQIVVTHDLQFABSISDTILKIN 207

```

-1696-

E GWT ++VTH++ FAE ++ ++ I+  
 Sbjct: 181 DLAEKGMTVMIVTHETGFAEKVASKIIFID 210

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4703> which encodes the amino acid sequence <SEQ ID 4704>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2146 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 164/209 (78%), Positives = 183/209 (87%)  
 Query: 1 MLEKNISKCYGQKEIFADFNLTVEBGKILSLVGPSSGGKTTLLRMAGLEKIDSGTIVH 60  
 MLEKNISK +GQK IF FNLTV+G++LSLVGPS GKKTTLLRMAGLE IDSG + +  
 Sbjct: 1 MLEKNISKQFGQKTTIDGFFNLTVQDGEVLSLVGPSGGKTTLLRMAGLESIDSGQVFF 60  
 Query: 61 DGKEVSVHDLETLNLLGFFVQDFQLFPHLTIVLNLILSPVNTWGLSKELAKEKALVLLER 120  
 +G++V +DHL E NLLGFFVQDFQLFPHLTIVLNL LSP TWG K AREKAL LL R  
 Sbjct: 61 NGEDVGDHLENRNLLGFFVQDFQLFPHLTIVLNLTLSPITTWGKKQDAKEKALDILLAR 120  
 Query: 121 LGLKDHAIYVFP+SLSGSQKQFVALARAMMIDPQIIIGYDEPTSAIDPELRQVEKLILQNR 180  
 LGLK+HA YVP+SLSGSQKQFVALARAMMIDPQIIIGYDEPTSAIDPELRQ VE LI+QNR  
 Sbjct: 121 LGLKERAQYVTP+SLSGSQKQFVALARAMMIDPQIIIGYDEPTSAIDPELRQVEALIVQNR 180  
 Query: 181 ETGMITQIVVTHDLQFAESISDITLKINFK 209  
 E G+TQIVVTHDL FAE+ISD I++NFK  
 Sbjct: 181 EHGITQIVVTHDLVFAEALSDRIIRVNFK 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 35 Example 1528

A DNA sequence (GBSx1619) was identified in *S.agalactiae* <SEQ ID 4705> which encodes the amino acid sequence <SEQ ID 4706>. This protein is predicted to be amino acid ABC transporter, permease protein (glnP). Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -8.12 Transmembrane 102 - 118 ( 96 - 120)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4248 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9341> which encodes amino acid sequence <SEQ ID 9342> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA98402 GBS:AP002545 ABC amino acid transporter permease  
 [Chlamydomonas pneumoniae J138]  
 Identities = 55/127 (43%), Positives = 83/127 (65%), Gaps = 1/127 (0%)  
 Query: 3 AAILATFMNYAAYFAEIPRGIESIPKGYEAAKVLKFSKFTQWYIVLPQVFKVLPSV 62  
 A IIA +MN AAY AE RGGI S+ GQ+E+A VL + K+Q YI+ PQVFK +LP8+

-1697-

Sbjct: 89 AGIIALSNNSAAYLAENIRGGINSISIGQWESAMVLGYKYQIFVYIYPQVFNILPSL 148

Query: 63 FNETITLVKDSILVYLGVGDLLESKTAANRDATLAPMF-TAGGIYILLIGLITILSKQ 121  
NE ++L+K+SS++ ++GV +L +K +R+ M+ I G+Y L+ + +S+

Sbjct: 149 TNEFVSLIKSSIIIMVVGVPFLTKVTKDIVSREINPMRYMLICAGLYPLMTSSFSICISRL 208

Query: 122 VIKRFPNY 128

EIR +Y

Sbjct: 209 SEKRFSY 215

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4707> which encodes the amino acid sequence <SEQ ID 4708>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

15	INTEGRAL	Likelihood	--11.57	Transmembrane	21 - 37 ( 7 - 44)
	INTEGRAL	Likelihood	--10.93	Transmembrane	185 - 201 ( 178 - 206)
	INTEGRAL	Likelihood	--3.29	Transmembrane	63 - 79 ( 62 - 81)

----- Final Results -----

20 bacterial membrane --- Certainty=0.5628(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:BAB05181 GB:AP001512 ABC transporter (permease) [Bacillus halodurans]  
Identities = 86/206 (41%), Positives = 126/206 (60%), Gaps = 1/206 (0%)

Query: 4 IQQVLPSLLDGLAVTLQVFFVIVIIISIPLAGIILAFIMKIPFKPLQNFILITLVVMRGSTPL 63  
IQ +P +L+G VTLQ ++ ++ +LG +LA ++ +NF Y + RGHPL

30 Sbjct: 8 IQPPMPFMLEGVVTLQFVSVLLPGLDVLGIVLAIKISKIRLFRNPFADFTYSIFRGITPL 67

Query: 64 LLQLIFFYYVLPSPGISFDRMEPAAILAFTLTAAYPAEIPFGGIEAIPKQYEAARKVLKL 123  
+LQL+ Y LP G+ + AA LAF LN AAY +RI R GI+ +KQ EA+ L +

35 Sbjct: 68 ILQLMIYIALPQFQVDLSIQQAFLAFLGNSAAYVSEIRAGIQAVDKGQREAEALGI 127

Query: 124 KPLQTIRYIILPQNVKIVLPSVFNENVLKDSSIVYVLGVGDL-LASKTAANRDATLA 182  
+ IILFQ + +LP++FNE INL K+S+V V+GV DL+ A TAA L

Sbjct: 128 FYRPMIRIILPQAMRNILPALFNEFINLTKESAIVSVIGVTDLMRRQAITSATYLYLE 187

40 Query: 183 PMFIAGLIYILLIGLIVTIISKQVEIR 208

P+ GLIY +L+ +T+I + +E+R

Sbjct: 188 PILFVGLIYVIVWGLTVIGRLERR 213

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 112/130 (86%), Positives = 121/130 (92%)

Query: 1 MPAAIIAFTMNYAAYFASIFRGGIESIPKQYEAARKVLKSKFQTVWYIVLPQVFKIVLP 60

MPAAI+APT+NYAAYFASIFRGGIE+IPKQYEAARKVLK QT+ YI+LPQVFKIVLP

Sbjct: 84 MPAAIIAFTMNYAAYFASIFRGGIEAIPKQYEAARKVLKAPLQTIKYIILPQVFKIVLP 143

50 Query: 61 SVFNETITLVKDSILVYLGVGDLLESKTAANRDATLAPMFAGGIYILLIGLITILSK 120

SVFNE I LVKDSILVY+LGVGDL L SKTAANRDATLAPMFAG TYLLIGL+TI+SK

Sbjct: 144 SVFNEVINLVKDSILVYLGVGDL L ASKTAANRDATLAPMFAGIYILLIGLIVTIISK 203

55 Query: 121 QVSKRFPNYK 130

QVSKRFPNY+

Sbjct: 204 QVSKRFPNYQ 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

60

-1698-

**Example 1529**

A DNA sequence (GBSx1620) was identified in *S. agalactiae* <SEQ ID 4709> which encodes the amino acid sequence <SEQ ID 4710>. This protein is predicted to be minidiscs. Analysis of this protein sequence reveals the following:

```

5    Possible site: 61
    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -9.66    Transmembrane    44 - 60 ( 39 - 66)
        INTEGRAL    Likelihood = -7.96    Transmembrane    129 - 145 ( 123 - 147)
10   INTEGRAL    Likelihood = -5.15    Transmembrane    13 - 29 ( 9 - 33)
        INTEGRAL    Likelihood = -2.39    Transmembrane    94 - 110 ( 94 - 110)

    ----- Final Results -----
        bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF49688 GB:AE003532 mnd gene product [Drosophila melanogaster]
Identities = 48/145 (33%), Positives = 78/145 (53%), Gaps = 8/145 (5%)

20   Query: 7    IKQTYGLMTTIAMIVGVVGGSGIYFKVDDILEFTGGDVFGLGMVILVGSFSIVFGSLSIS 66
        +K+ GL+ +A+IVGV++GSGI+ +LKT+ G + +++ VL + G+L +
    Sbjct: 39    LKKIQLGLDGVVAILVGVVGGSGIFVSPKGVWKS- GSGIGSLIWWLGSVLSMVGALCYA 97

25   Query: 67    ELAIRTSESGGIFSYYEKYVSPALAATLGLFASFLYL-PTLTAIVSVVAAPYTLGE---- 121
        EL +SGG ++Y F L A L L+ + L L FT AI + A Y L
    Sbjct: 98    ELGTMIPKSGGQYAYIGTAFGP-LPAFLYLMVALLILVFTGNATLTAFATYLLKPFWPS 156

    Query: 122   -SSSLESQIILAAYVILALSLMNIF 145
        + +S+ +LAA I L+L+N +
30   Sbjct: 157   CDAPTEAVQLLAAMICVLTILNICY 181
  
```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1530**

A DNA sequence (GBSx1621) was identified in *S. agalactiae* <SEQ ID 4711> which encodes the amino acid sequence <SEQ ID 4712>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 21
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
45   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
  
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1699-

## Example 1531

A DNA sequence (GBSx1622) was identified in *S. agalactiae* <SEQ ID 4713> which encodes the amino acid sequence <SEQ ID 4714>. This protein is predicted to be TRK potassium uptake system protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 27
   >>> Seems to have no N-terminal signal sequence
       INTEGRAL    Likelihood = -0.06    Transmembrane 232 - 248 ( 232 - 248)

10  ----- Final Results -----
       bacterial membrane --- Certainty=0.1022 (Affirmative) < succ>
       bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8835> which encodes amino acid sequence <SEQ ID 8836>
15 was also identified. Analysis of this protein sequence reveals the following:

       Lipop: Possible site: -1    Crend: 5
       MG: Discrim Score: -4.65
       GVH: Signal Score (-7.5): -3.64
       Possible site: 27
       >>> Seems to have no N-terminal signal sequence
       ALOM program count: 1 value: -0.06 threshold: 0.0
           INTEGRAL    Likelihood = -0.06    Transmembrane 228 - 244 ( 228 - 244)
           PERIPHERAL Likelihood = 1.27    428
       modified ALOM score: 0.51

25  *** Reasoning Step: 3

       ----- Final Results -----
       bacterial membrane --- Certainty=0.1022 (Affirmative) < succ>
       bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

35  >GP:AAB90401 GB:AB001046 TRK potassium uptake system protein
       (trkA-2) [Archaeoglobus fulgidus]
       Identities = 136/446 (30%), Positives = 238/446 (52%), Gaps = 12/446 (2%)

Query: 5   MRILVGGGKVGATALCRSLVAEKHDVVLIERKENVLKRVTKQHDIMQIVNGNANYKILEQ 64
       MRI++ G G+VG L SL A HDV++IRK + +RV++ D++ I GN AN K+L
40  Sbjct: 1   MRIVAGAGEVGYHLAMSL-APNDHVIILIKDVSRFERSSEL-DVAINGNAANNKVLRLD 58
       MRVKNCDIFIAITDRDEVNMISAVLAKKMGAKETVVRMRNPEYSNPFYKDKNPLGFSVV 124
       A V+ D+F+A+T DEVN++S + AKK+GAK +VR+ NPEY + ++ LG+ ++
45  Sbjct: 59  AGVERADVFLAVTGNDEVNLLSGLAAKKVGAKNIVVRNPEYVDRPIVKHPLGYDVLII 118
       NPELLAAQYIANTIEFPNATSVEHFAKGRVIMMRPKILGNKLCHTSMSPQIRKPKGNIVI 184
       P+L AQ A I P A V F+ G+V ++E ++EG+K +++ + N+VI
50  Sbjct: 119 CPQLSLAQEARLIGIPGAIEVVTFSGGKVMIEIQVMSGSKADKAIDLYLP-QNVVI 177
       CATRDGKLII PDGDATIQVDDKI FVTGNRIEMILFHINVRKNVKNLAVIGAGRIAYYL 244
       +I R+G + IP GD ++ D++ + ++ V + + + GAG I Y
55  Sbjct: 178 ASIYRNGHIEIPGDTVLRAGRDVAIVS KTDVEMLKGI FGPVPVTRRTVI FPGATIGSYT 237
       INILANTNTHVKLVELNQBQAEYFSQEPNVPVHVGDTAKNILLIESVTSFDAVATLTG 304
       IL T VKL+E + E+ S E S V +V GD T L+ES + DAV T
60  Sbjct: 238 AKILAKMGFVSVKLIRSSMERCKALSGELGVRTVCGDATDIEFLIERETIGKSDAVLAATE 297
       VDEENITSMFLESIGIPKNITKVNRKTSLEIIDDQLSSIIITPKRIADVHMHFVRGRV 364
       DHN++ S+ +++G I KV + ++ + + + D+ + + V +R
65  Sbjct: 298 SDEKNLLISLLSKNLGARIAIAKVKREYVYKLFPAVGVDVALNFRSVTYNRSKLR--- 354
       NAQSNLEAMPHIANDRIETLQFIKETSILANRSLASLKLQNILIAAIRNNKTIPT 424

```

-1700-

+E + I + + + ++L ++L L L ++ +I AI+R N+ + P  
 Sbjct: 355 ---TMRITLAEIGTAVVEV---VVRNTRLVGKALKDLPLPKDAIIGAIVRGNECLIPR 408  
 Query: 425 GEDVLTVGDRIVVITLKNITRTSD 450  
 5 G+ + DR++V I + ++  
 Sbjct: 409 GUTTIETEDRLIVFAKMDIEIKIEI 434  
 Identities = 48/212 (22%), Positives = 99/212 (46%), Gaps = 15/212 (7%)  
 Query: 3 VQRITIVGGKQVTLARSLVAEKHDVVLIEKKENVLKRVTKQHDIMHIV-NGNANYKI 61  
 10 V R+ + G G +G+ + L V LIE + + + + + IV G+ + +  
 Sbjct: 221 VTRRVTTFGAGTIGSYTAKILAKMTSVKLISSMERCEALSGELGVVRIVGSDATDIEF 280  
 Query: 62 LEQAQVKNCDIFIAITDRDEVMISAVLAKMKAKETVVRMRNPETSNPYFKDKNPLGFS 121  
 L + E+ D +A T+ DE N++ ++L+K +G+ + + + EY + +G  
 15 Sbjct: 281 LKKEEIGKSDAVLAATESDEKNLLISLLSKMLGARIAIAKVEKREYVKLF----EAVGVD 336  
 Query: 122 SVNPPELLAAQYIA---NTIEFFNATSVHEFANGRVPLA/EPKILEGNKLCHTSMQIRKK 178  
 +NP + + + T+ +E A V++ +++ G L + +  
 20 Sbjct: 337 VALNPRSVTYNEVSKLLRTMRITLAEIGTAVVEVVVRNTRLV-GKALKDLPLPK---- 391  
 Query: 179 FGNIVICAIERDGLIIPGDATIQVQDKIFV 210  
 + + I AI R + +IP GD TI+ +D+ + V  
 Sbjct: 392 --DAIIGAIVRGNECLIPRGDTTIEYEDRLIV 421

25 There is also homology to SEQ ID 4716.

SEQ ID 8836 (GBS384) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 2; MW 53kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 6; MW 78kDa).

30 The GBS384-GST fusion product was purified (Figure 212, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 279), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1532

35 A DNA sequence (GBSx1623) was identified in *S. agalactiae* <SEQ ID 4717> which encodes the amino acid sequence <SEQ ID 4718>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence  
 40 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.4948 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1701-

**Example 1533**

A DNA sequence (GBSx1624) was identified in *S. agalactiae* <SEQ ID 4719> which encodes the amino acid sequence <SEQ ID 4720>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
  INTEGRAL    Likelihood =-12.58    Transmembrane    37 - 53 ( 33 - 61)
  INTEGRAL    Likelihood =-11.57    Transmembrane    183 - 199 ( 179 - 214)
  INTEGRAL    Likelihood =-10.03    Transmembrane    397 - 413 ( 392 - 424)
  INTEGRAL    Likelihood = -6.79    Transmembrane    14 - 30 ( 5 - 31)
10  INTEGRAL    Likelihood = -6.42    Transmembrane    71 - 87 ( 69 - 93)
  INTEGRAL    Likelihood = -4.99    Transmembrane    278 - 294 ( 274 - 295)
  INTEGRAL    Likelihood = -4.19    Transmembrane    133 - 149 ( 132 - 152)
  INTEGRAL    Likelihood = -4.09    Transmembrane    327 - 343 ( 324 - 344)
  INTEGRAL    Likelihood = -2.44    Transmembrane    236 - 252 ( 234 - 252)
15  INTEGRAL    Likelihood = -0.59    Transmembrane    456 - 472 ( 456 - 472)

----- Final Results -----
      bacterial membrane --- Certainty=0.6031 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
20  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10065> which encodes amino acid sequence <SEQ ID 10066> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

25  >GP:AAB90400 GB:AE001046 TRK potassium uptake system protein (trkH)
    [Archaeoglobus fulgidus]
    Identities = 166/480 (34%), Positives = 262/480 (54%), Gaps = 10/480 (2%)

30  Query: 1 MNKSIRFLLSKILLTEAALLAIPITVGLIYREP-QSVMMISIVITMIILILGLGLSLFK 59
    MN + +L KLL++ + +FL ++ EP ++ +++++ +LG G +
    Sbjct: 1 MNLRLTASILKILMLFSFSFILPLLAHVPEEPYHPFLIPALSLILGAVLGV-GIKTE 59

    Query: 60 PKNYHIYTKGMLIVALCWILMSFFGALPFVISGQIPNIIDAFPEVSSGPTTTGATILDD 119
    + + KE IVAL W+ S FG++P++I G P +DAFFE SGPTTTGA++L
35  Sbjct: 60 SEFDSLHKESFAIVALIWLFWMSIRGSPYIIFGISP--VDAFFESMSGPTTTGASVITP 117

    Query: 120 VSVLSPALLFWRSPTHLIGGMKVLVFAIAIMENSKNSHLEVMAEVPGVFGKVVSKLKK 179
    L +LL WRS T IGGMG++V LAI N + +AE PG K+ +++
40  Sbjct: 118 EE-LPKSLLWRSILQWIGMGIIIVLFLAIFPNVAKRSTVLFGAEYPGVSLSKLKPRI 176

    Query: 180 TAQILYLILYLMFAVFAVILYFAGMPFFDSIIAMOTAGTGGFAVYNDIAHYSPLITN 239
    TA LY +YLL+ +LY G+ PD+I T TGG++ ++SIA + +
50  Sbjct: 177 TALSILKYVLLLTARVALLYALGLSLFDAINHTFTTLTGSGYSTHSEIAFFKDVVEA 236

    Query: 240 LVSGIMLIPGVNFNLYLLLRKIKAFPGDEELKTYLRIVAIATPMIAINWIGNYNPRO 299
    YV+ + G NF L Y LL K F + E + Y+ +A+A+ +IA + Y F +
55  Sbjct: 237 VVAFAPLGGNFALITYLLSCK-PVIFRNTFRAYVCFPLASVIAAVNLDRYSIF-E 294

    Query: 300 GLEHIFFEVSAILTTGFGVTDITRWPLFSQVILLFLMFIGSAGSAGSGFKVNRSLILA 359
    L + F+ +I+TTTGF D W +++LL+ IMFIOGS+GST GG KV+R +L
60  Sbjct: 295 SLRYSIFQVNSIMTTTGFTADFDAWDSKSLILVVMFPIGSSGSGTGGGIKIRIYLLI 354

    Query: 360 KIARNOVLSTLYPNKRVMSHLINKSVLDKNTQBGVLKYLITYLAI PMALVILVLTLDNDFL 419
    KA +Q+L P V ++ + K + + +Y+ IF ++++L D +
55  Sbjct: 355 KYAVHQLRAAEFRTVRVAKFEGRAKKKILDDIAAFPVYIILI FAVSSILVLSGYDIV 414

    Query: 420 VVISAAACFNNGP---LLGSNETFSFFSPSKILLSPANIAGRLEIYVPLMFIPIKTW 476
    ISA A+ N+GP L G+ E ++ F +K+LL+ M GRLEI+ V+ +FIP W
60  Sbjct: 415 TSISATAATLGNVGRGLGAGAEVYASFPSTKILLAVNMWIGRLEIFTVSVLSPTFW 474

```

No corresponding DNA sequence was identified in *S. pyogenes*.

-1702-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1534

A DNA sequence (GBSx1625) was identified in *S. agalactiae* <SEQ ID 4721> which encodes the amino acid sequence <SEQ ID 4722>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence (or aa 1-20)

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2870 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD36530 GB:AB001797 conserved hypothetical protein
[Thermotoga maritima]
Identities = 43/75 (57%), Positives = 57/75 (75%), Gaps = 1/75 (1%)

Query: 2 LKSEFLIFLVRFYQKNISPAFFASCRYPPTCSYTMIEAIQKHG-LKGVLMGIARILRCHPL 60
      +K LI L-RFYQ+ ISF F +CR+ PTCS Y I+A++KHG LRG +G+ RILRC+FL
Sbjct: 1 MKKLLILIRFYQRYISPLKPPPTCRFTPTCSNYFIQALEKHGELLGTFPLGLRILRCNPL 60

Query: 61 AHGQNDVFDPHPSLR 75
      + GS DFPV+ FS +
Sbjct: 61 SKGYDVFPEEBSFK 75
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4723> which encodes the amino acid sequence <SEQ ID 4724>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3639 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 53/78 (67%), Positives = 60/78 (75%)

Query: 1 MLKSEFLIFLVRFYQKNISPAFFASCRYPPTCSYTMIEAIQKHG-LKGVLMGIARILRCHPL 60
      M+K LI V+ YQK ISF P SCRY+PTCS YM+ AI+KHG KG+LMGIARILRCHP
Sbjct: 1 MMKLLIVSVKAYQKYISPLSPSCRYKPTCSAYMLTAIEKHGTFGLMGIARILRCHPF 60

Query: 61 AHGQNDVFDPHPSLRNK 78
      GG DFPV+ PSL RNK
Sbjct: 61 VAGGVDFVDEDSLRNKK 78
```

SEQ ID 4722 (GBS233) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 3; MW 35.6kDa).

The GBS233-GST fusion product was purified (Figure 207, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 280), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.



-1703-

**Example 1535**

A DNA sequence (GBSx1626) was identified in *S. agalactiae* <SEQ ID 4725> which encodes the amino acid sequence <SEQ ID 4726>. This protein is predicted to be ribosomal large subunit pseudouridine synthase B (ruB). Analysis of this protein sequence reveals the following:

5      Possible site: 18  
      >>> Seems to have no N-terminal signal sequence

10      ----- Final Results -----  
          bacterial cytoplasm --- Certainty=0.2957 (Affirmative) < succ>  
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15      >GP:BAB05295 GB:AP001512 pseudouridylate synthase [Bacillus halodurans]  
      Identities = 130/239 (54%), Positives = 175/239 (72%), Gaps = 2/239 (0%)

Query: 2    RINKYIAHAGIASRKAELIKQGMVITINGQVVELATQVKAG-DLVEIEGSPYINBEKV 60  
      R+ K IA AGIASRKAELI +G V +NGQVV EL +V    D +E+EG P+ EK V  
      Sbjct: 3    RLQKVIAQAGIASRKAELIKGLGKRVKVGQVVKELGKIVNQDDIEVEGVFVEKEEPV 62

20      Query: 61    YYLLNKPRGVISSVSDDKGRKTVIDLPLQVKERTYFPVGRLDWDTGLLLITNDGDFPTDM 120  
      Y+LL KP GVISSV DDKGRK V D L +++R+YFVGRLD+DT+GLL+LITNDG+P + +  
      Sbjct: 63    YFLLYKPTGVISSVSDDKGRKVVTDPL-EIQRVYFPVGRLDYDTSGLLLLITNDGEPANLL 121

25      Query: 121    IHPNEIDKIVLARVGIATKENLRPLTRGVVIDGKKTTPARYTIIKVDHEKNSRVVELT 180  
      +HFR++I+KVY+A+VIGL T++ L+ L RGV ++    T PA+ ++ VD K ++V+LT  
      Sbjct: 122    MHPRHKIEKVVAKVIGLPTRDQLKLLARGVKLEGGTAPAKVQMLSVDRRKQTAVKLT 181

30      Query: 181    IHEGRNHQVKMPQVGLLDVKLSRTQFGTLDLTGLRPGEARLNKKESQLHNAAINK 239  
      IHEGRN QV++MFE +G V KL R OF LDL+G+ PG+ R L E+ L A+ K  
      Sbjct: 182    IHEGRNHQVEMFETIGCEVMKLEQVFAFLDLGSMNPGDVRLKPIEVKHLRLAIVTK 240

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4727> which encodes the amino acid sequence <SEQ ID 4728>. Analysis of this protein sequence reveals the following:

35      Possible site: 18  
      >>> Seems to have no N-terminal signal sequence

40      ----- Final Results -----  
          bacterial cytoplasm --- Certainty=0.1587 (Affirmative) < succ>  
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45      Identities = 210/239 (87%), Positives = 228/239 (94%)

Query: 1    MRINKYIAHAGIASRKAELIKQGMVITINGQVVELATQVKAGDLVEIEGSPYINBEKV 60  
      MRINKYIAHAGIASRKAELIKQGMVITINGQVVELATQVKAGDLVEIEGSPYINBEKV  
      Sbjct: 9    MRINKYIAHAGIASRKAELIKGLVITINGQVVELATQVKAGDLVEIEGSPYINBEKV 68

50      Query: 61    YYLLNKPRGVISSVSDDKGRKTVIDLPLQVKERTYFPVGRLDWDTGLLLITNDGDFPTDM 120  
      YYLLNKPRG ISSVSDDKGRKTV+DLLPQVKERTYFPVGRLDWDT+G+LITNDGDFPT M  
      Sbjct: 69    YYLLNKPRGALISSVSDDKGRKTVIDLPLQVKERTYFPVGRLDWDTSGVLLITNDGDFPTM 128

55      Query: 121    IHPNEIDKIVLARVGIATKENLRPLTRGVVIDGKKTTPARYTIIKVDHEKNSRVVELT 180  
      IHPNEIDKIVLARVGIATKENLRPLTRGVVIDGKKTTPARYTIIKVDHEKNSRVVELT  
      Sbjct: 129    IHPNEIDKIVLARVGIATKENLRPLTRGVVIDGKKTTPARYTIIKVDHEKNSRVVELT 188

60      Query: 181    IHEGRNHQVKMPQVGLLDVKLSRTQFGTLDLTGLRPGEARLNKKESQLHNAAINK 239  
      IHEGRNHQVKMPQVGLLDVKLSRTQFGTLDLTGLRPGEARLNKKESQLHNAAINK  
      Sbjct: 189    IHEGRNHQVKMPQVGLLDVKLSRTQFGTLDLTGLRPGEARLNKKESQLHNAAINK 247

-1704-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1536**

A DNA sequence (GBSx1627) was identified in *S. agalactiae* <SEQ ID 4729> which encodes the amino acid sequence <SEQ ID 4730>. Analysis of this protein sequence reveals the following:

Possible site: 15  
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1476 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA05280 GB:AP001512 unknown conserved protein [Bacillus halodurans]
Identities = 75/180 (41%), Positives = 107/180 (58%), Gaps = 10/180 (5%)

Query: 6  SIEALLFVAGEDGLSRQMAELLSTPPSALIQQLEKLAKRYEEDDDSSLLLETAQTYKL 65
+IE +LFV G+G+L ++ +LL L+ + LE+L Y D+ L + E A ++L
Sbjct: 9  AIEGILFVAGEDGVLEELCDLLELSTVVYAAELRLSYT-DEAGLRIRIEVAHAFRL 67

Query: 66  VTQDSYMTLLRDYAKAPINQSLSRASLEVLSTIAYKQPTITRIEIDDIRGVNSSGAITRLI 125
TK + A + + LS+A+LE L+IAY+QPTIRIE+D+RGV S AI L
Sbjct: 68  STKPELAYPFKKLALSTLQSGLSQAALSTLAIAYRQPTITRIEIVDEVGRVNSEKAIQTLL 127

Query: 126  AFGLIKKEAGKKEVLGRPNLYETTYFLDVMGINQLDLDL-----IDASSIELVDEEVSFL 179
+ LIKE G+ + GRP LY TT FLD+ G+ L +L ID SSI EE LF
Sbjct: 128  SRLLIKEVGRAQGTGRPILYGTTTQFLDHFGLKSLKSLPLPEDIDESSI---GEEDLF 184

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4731> which encodes the amino acid sequence <SEQ ID 4732>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1062 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 130/179 (72%), Positives = 159/179 (88%)

Query: 1  MYYLSGIEALLFVAGEDGLSRQMAELLSTPPSALIQQLEKLAKRYEEDDDSSLLLETA 60
MYTL IEALLFVAGE+GLSR +A +LSLTP+AL QQLEK+++YE+D SSL L+ETA
Sbjct: 1  MYYLSQIEALLFVAGEGLSIRHLSNLSLTPPALQQLEKLSQYKQDKQSSICLIETA 60

Query: 61  QTYKLVTQDSYMTLLRDYAKAPINQSLSRASLEVLSTIAYKQPTITRIEIDDIRGVNSSGA 120
TY+LVTK+ + LLR YAK P+NQSLSRASLEVLST+AYKQPTITRIEIDDIRGVNSSGA
Sbjct: 61  NYYRLVTKSGFAKLLRAYAKTPMNQSLSRASLEVLSTIVAYKQPTITRIEIDDIRGVNSSGA 120

Query: 121  ITRLIAFGLIKEAGKKEVLGRPNLYETTYFLDVMGINQLDLDASSIELVDEEVSFL 179
++L+AP LI+EAGKK+V+GRP+LY TT+YFLDVMGIN LD+LI+ S+E DEE+LF
Sbjct: 121  LSKLLAPDLIREAGKKDVVGRPNLYATTDTYFLDVMGINHLDLIEVSGEVADEESIALF 179

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1705-

**Example 1537**

A DNA sequence (GBSx1628) was identified in *S. agalactiae* <SEQ ID 4733> which encodes the amino acid sequence <SEQ ID 4734>. Analysis of this protein sequence reveals the following:

Possible site: 47  
 >> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1012 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 89/246 (36%), Positives = 145/246 (59%), Gaps = 19/246 (7%)  
 Query: 3 IKLKDPEGFLDLLLHLVSKVEVDIYDVEIVEVIEQYLAYIATLQMRLEVAGEYMLMASQ 62  
           +K+ FEGFLDLLLHL+++ E+DIY+P+ ++ EQYL Y+ T++ L++A EY++MA+  
 Sbjct: 6 VKIDTFEGFLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65  
 Query: 63 LMLIKSRNLLPK-----VVESNPI-EDDPMEILLSQLEEYRFRKVLSEELANGHOERAKYF 117  
           L+ IKSRLLPK + E + E+DP EL+ +L EYR++K +++L + +ER K F  
 Sbjct: 66 LLSIKSRMLLPKQEEELFEDELLLEEDPREELIEKLEIYRKYDNDKDKEREERQKSF 125  
 Query: 118 SKPKQEVTFEDAILLHDKSVMDLFLTFSCMMSQKQKELNS-----QTVIEKEDYRIED 171  
           +KP ++ + +S L +T M+ OK L +T I ++D IE  
 Sbjct: 126 TKPPSDL--SEYAKVQSQKLSVTVVYDMIGAPQKVLKRCINRPMETITRODPIEA 183  
 Query: 172 MMIVIERHFNLLKKTT--LQEVFADCTKSEMITLFLAMLELIKHLQITVQDSNFSQV 228  
           M I +LK +T ++F + K ++ FLA+LEL+K + +BQ+ NFG +  
 Sbjct: 184 RMNEIVH--SLKSRGTRINFMDLF-PYEQKEHLVVTFLAVLELMKQQLVLSIQEHNFSDI 240  
 Query: 229 ILRKEE 234  
           + E  
 Sbjct: 241 YITGSE 246

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4735> which encodes the amino acid sequence <SEQ ID 4736>. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -3.61 Transmembrane 199 - 215 ( 199 - 218)  
 ----- Final Results -----  
           bacterial membrane --- Certainty=0.2444 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 86/239 (35%), Positives = 145/239 (59%), Gaps = 15/239 (6%)  
 Query: 3 IKLKDPEGFLDLLLHLVSKYKVDIYEVPIVEIVEQYINYYIETLQVMKLEVAGDYMLMASQ 62  
           +K+ FEGFLDLLLHL+++ ++DIY++P+ ++ EQYL Y+ T++V++L++A +Y++MA+  
 Sbjct: 6 VKIDTFEGFLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65  
 Query: 63 LMLIKSRRLLPKVEHI-----BEDLEQDLLEKIEEYRFEAVSQALAKHODQRAKY 115  
           L+ IKSRLLPK E + BED ++L+EK+ EY ++K ++ L ++ ++R K +  
 Sbjct: 66 LLSIKSRMLLPKQEEELFEDELLLEEDPREELIEKLEIYRKYDNDKDKEREERQKSF 125  
 Query: 116 SKPKQELI-FEDAILQEDK---TVMDLFLAPSINMAKRAVLKNNHVTIERDYDIKEDM 170  
           +KP +L + + Q ++ TV D+ AF ++ K+ ++ T I R D IE  
 Sbjct: 126 TKPPSDLSEYAKVQSQKLSVTVVYDMIGAPQKVLKRRK-INRPMETITRODPIEAR 184

-1706-

Query: 171 MASIKQRLKEINV-IRLSAIFECQTLNEVISIFLASLELIKLVVVFVQLSNFGAILL 228  
 M I L+ I +F Q + V++ FLA LRL+K +V +RQ NF I +  
 Sbjet: 185 MNEIVHSLKSRGTRINFMDLFPYQKHLAVVT-FLAVLELMKNQVLRLQKHNFDIYI 242

An alignment of the GAS and GBS proteins is shown below.

Identities = 156/235 (66%), Positives = 191/235 (80%), Gaps = 2/235 (0%)

Query: 1 MDIKLDFEGFDLILLHLVSKYEVDIYVPVIVIEQYIATLQMRLEVAGSYMLNA 60  
 MDIKLDFEGFDLILLHLV8+Y+VDIY+VPVIVIEQYI YI TLQ N+LEVAG+YMLNA  
 Sbjet: 1 MDIKLDFEGFDLILLHLV8QKVDIYEVPIVIEQYIYIETLQVMKLEVAGDYMLNA 60

Query: 61 SQMLIKSRNLLPKVVENPIKDDPEMELLQLEIYHRFKVISRELANQHQBRKYPSPK 120  
 SQMLIKSR LLLPKVVE IS+D E +LL ++EY RPK +S+ LA QH +RAK+SKP  
 Sbjet: 61 SQMLIKSRLLLPKVVH--TEEDLEQDILLKIEYSRFRKAVSQALAKQHDQRAKYSK 118

Query: 121 KQEVIFEDAILLHDKSVMDLFTFSQMMSQKQKELNSQTVIEKEDYRIEDNMIVIERHF 180  
 KQEVIFEDAIL DK+VMDLFL FS +M+ K+ L N+ TVIE++DY+IEDNM I++  
 Sbjet: 119 KQELIFEDAILQBDKTVMDLFLAFSNIMAAKRAVLKNNHTVIERDDYKIEDNMASIKQRL 178

Query: 181 NLKKTTLQEVFADQCTSEMITLFLAMELELIKHLQITVEQDSNFSQVILRKEE 235  
 + L +F +CQT +E+I++FLA LELIKH + VEQ SNF +ILRKE+K  
 Sbjet: 179 EKENVIRLSAIFECQTLNVSISIFLASLKLILKLVVVFVQLSNFGAILRKEE 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1538

A DNA sequence (GBSx1629) was identified in *S. agalactiae* <SEQ ID 4737> which encodes the amino acid sequence <SEQ ID 4738>. This protein is predicted to be pXO1-18. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.14 Transmembrane 128 - 144 ( 127 - 145)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2657 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05248 GB:AP001512 integrase/recombinase [Bacillus halodurans]  
 Identities = 67/271 (24%), Positives = 117/271 (42%), Gaps = 35/271 (12%)

Query: 11 LRTMINDINNFIKSK----LSLSRKSRYHYDLKQFYKII-----GCHVNSKALAY 58  
 ++T+ N++ F+ +K LS N++SY DLQ+ + + ++ E + Y  
 Sbjet: 1 METVNNLQQFLHFPQKVERGLSNNTIQS YGRDLKQYIQYQVHVREIRARNITRETIHY 60

Query: 59 QQSLSKFL--TARKEKLSAVNQFLFLVNRGTLKEFYRL-----QSTHKITLAQTSQI 111  
 L E T+ R +R+ F FL + + T++ A T ++  
 Sbjet: 61 LVHLRQGRNASTIARAAIRSPHQFLRLREKLSDSDPTVHVPEIKATKRLPALATIEEV 120

Query: 112 MDLSNFYQDTYPSGRLTALLIL--SLGLTPAEIANLKKADFTTNILS-IEKSMKRI 168  
 L N Q D S R A+L L +G+ +R+ L +D + + + K +RI  
 Sbjet: 121 EALLNSPQGRDPPSLNKAEMELLVATGMRSVLEIGLTSLDTHLSMGVPRCIRKGNKRI 180

Query: 169 LKLPEDLLPFLLESLEEDG-----DLVF-IEKNGPYSRQWFFNQIDFLINEN-R 216  
 + + + + +ES +G D VF H+G+P SRQ F+ L N +  
 Sbjet: 181 IPIQV-VATEAVESYLANGRGKLMKKQSHDEIVFVNHHRPLSRQGWMLKQLAKNVND 239

Query: 217 QQTUQLLEQFILKQKNGKNTIMLSRLGL 247

-1707-

+ LT LR F ENG + + +LG  
 Sbjct: 240 KPIPTHTLRHSFATHLLNGADLRVQBMIG 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4739> which encodes the amino acid sequence <SEQ ID 4740>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.90 Transmembrane 111 - 127 ( 110 - 127)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1362 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/243 (48%), Positives = 167/243 (68%), Gaps = 1/243 (0%)

Query: 18 INNPIESKKLSINRSKSYHYDLKQFYKIIGHVNSEKIALYQOQLSEPKLTARKRKLSAV 77  
 I FI SK LS NS+K+Y YDL+QF ++IG VN +KL LYQ S++ L+A+KRKLS  
 Sbjct: 5 IEPFIASHKLSQNSQKAYRYDLQFQQLIGERVMDKLLYQNSIANLSLSAKRKLSA 64

Query: 78 NQPLFFLYNRGTLEKEFYRLQETEKITLAQTK-SQIMOLSNFYQDTDYPSGRLLIALLLSL 136  
 NQPL+ +LY L ++RL +T K+ + + + I++ FYQ T + G+LI+LLIL L  
 Sbjct: 65 NQPLXYLYQIKYLSYFRLLDTMKVMRTEKQQAALINTDIFQKTPFVWQQLISLLILEL 124

Query: 137 GLTPARIANLKKADFDFTFNILSIEKSQMKRILKLPEDLLPFILESLEBQDGLVFEHNGK 196  
 GLTP+R+A ++ A+ D F +L+++ + R+L L + L+PFL + L +FEH G  
 Sbjct: 125 GLTPSEVAGIEVANLDLNFQMLTLTKKGVRLPLSQLIFPLBQQLVGEVLYFEHNGI 184

Query: 197 PYSRQMFNFQLTDFLANEKNQQLTAQLLREQQLKQENKIMTELSRLGLKTPITILR 256  
 P+SRQMFNF L F+ + LTAQ LREQFLK+K GK++ ELS +LGLK+P+TLE+  
 Sbjct: 185 PFSRQMFNFHLKTFVRSIGYBGLTAQKLREQFLKEKLAKGSIIEISDILGKSPMTLEK 244

Query: 257 YYR 259  
 YY+  
 Sbjct: 245 YYK 247

SEQ ID 4738 (GBS383) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 7; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 5; MW 57.1kDa).

The GBS383-GST fusion product was purified (Figure 212, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 308), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1539

A DNA sequence (GBSx1630) was identified in *S.galactiae* <SEQ ID 4741> which encodes the amino acid sequence <SEQ ID 4742>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2465 (Affirmative) < succ>

-1708-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BA05201 GB:AP001512 unknown conserved protein in *B. subtilis*  
 [Bacillus halodurans]  
 Identities = 38/136 (27%), Positives = 73/136 (52%), Gaps = 1/136 (0%)

10 Query: 7 ESFLNHLNHYLIPAEVVAIFVDTHNADRVMLLLASNGFSRVPVITREKKYVGTISISDI 66  
 ++ + N L +IP E VA ++ +H +L+L +G++ +PV+ + K G IS S I  
 Sbjct: 7 QNIMNDLKLIVPFKVAHVLSNPLEHALLVLKSGYTAIPVLDRHSLKHGVISKSIL 66

Query: 67 MGYSKQQLTWB-MAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISDQNDFR 125  
 + + + E +A + ++N +I I+ A+ + + +PF+ ++ D F  
 15 Sbjct: 67 LDALLGVRIEMERIAHLVVKDVMNPEIPTTHHKASFSRALKVSLAHPFCILDDGSGFL 126

Query: 126 GIITRKSILKAINSLG 141  
 GI+TR +IL IN L  
 Sbjct: 127 GILTRSTILSFINRQL 142

20

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4743> which encodes the amino acid sequence <SEQ ID 4744>. Analysis of this protein sequence reveals the following:

Possible site: 47  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3539 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 119/153 (77%), Positives = 137/153 (88%)

35 Query: 1 MIAKEFESEFLNHLNHYLIPAEVVAIFVDTHNADRVMLLLASNGFSRVPVITREKKYVGT 60  
 MIAKEFE+FL++HLD+YLIP +D+AIF+DTHNADRVMLLL SNGFSRVPVIT+EKYVGT  
 Sbjct: 1 MIAKEFETLMSHLNLYLPEQLAIFIDTHNADRVMLLLVSNFSRVPVITREKKYVGT 60

Query: 61 ISISDINGYSKQQLTWBMAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISD 120  
 ISISDIN YQSK QUTWEM+QTDI EMVNTKIE I+ ++LT IMHK++D+PFLPV+  
 40 Sbjct: 61 ISISDINMYQSKRQLTDWMSQTDIGEMVNTKIEITISITSLTEIMHKLIDFPFLPVDR 120

Query: 121 QNDFRGIITRKSILKAINSLHDFDTDEYITTPK 153  
 N F GIITRKSILKA+NSLLHDFDT+YTI K  
 45 Sbjct: 121 ANRFVGIITRKSILKAVNSLLHDFDTDYITTIKK 153

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1540

50 A DNA sequence (GBSx1631) was identified in *S. agalactiae* <SEQ ID 4745> which encodes the amino acid sequence <SEQ ID 4746>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 55 bacterial cytoplasm --- Certainty=0.4421 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1709-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06705 GB:AP001517 unknown conserved protein [Bacillus halodurans]  
Identities = 55/169 (32%), Positives = 95/169 (55%), Gaps = 1/169 (0%)

5 Query: 5 KLVVMSDSHGDRDIVDKIKHYLGKVDALPHNGDSELPSSDPIWEGIHVVTGNCIDYSGY 64  
KL++SDSHG D +K + + + +VDAI H GDSLEP D EG++V GNCID+ +  
Sbjct: 2 KLLILSDSHGWSDELKAVADKHQEVDAIITHGDSLEPDRDRALSGNIVRGNCDFGVDF 61

10 Query: 65 PEVLVTIKIDNAVIVQTHGHILQINFTWDLKDLAQEDADICLYGHILHRADAWNGKTIIF 124  
PE + + + + THGHIL+ + + L A++ A + +GH A + + + +F  
Sbjct: 62 PRDFIKTVGDPNVVYTHGHILYVNMGSYSLTYRABEVGALVCFGSHVATSPQSGNIVF 121

15 Query: 125 INFGSVLPQPRGPINEKLYAVVTITDSKVLVEYYTTRHQPIYNLTKELSR 173  
+NFGS+ PR E+ Y + + D ++ + + R +L+ + R  
Sbjct: 122 VNPQSLRLPRNR-KBOTYCLAYVRDDQIKELVFLDRDGHEVTDLQRTYLR 169

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4747> which encodes the amino acid sequence <SEQ ID 4748>. Analysis of this protein sequence reveals the following:

Possible site: 14  
20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3835 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/173 (67%), Positives = 143/173 (82%)

30 Query: 1 MAIRKLVVMSDSHGDRDIVDKIKHYLGKVDALPHNGDSELPSSDPIWEGIHVVTGNCIDY 60  
MA + ++VMSDSHGDRDIV+ IK+ YLG+VDAIPHNGDSEL SSDPIW GI+VV GNCIDY  
Sbjct: 1 MASTKIIVMSDSHGDRDIVQAIDKQYLGQVDAIPHNGDSELSNDSPIWAGIYVVGNCIDY 60

35 Query: 61 DSGYPEVLVTIKIDNAVIVQTHGHILQINFTWDLKDLAQEDADICLYGHILHRADAWNG 120  
D+GYP+ LVT++ I QTHGHIL+ INFTWDLKDL AQ+ ADICLYGHILHR AW+ G  
Sbjct: 61 DTGYEDRLVTLQGTVTIAQTHGHILYHINFTWDLKDYFAQEVVADICLYGHILHRPAWQVG 120

Query: 121 KTIIFNFGSVLPQPRGPINEKLYAVVTITDSKVLVEYYTTRHQPIYNLTKELSR 173  
+T+FNFGSV QPRG INEKLVA V +TD++ V+Y+TR H+ YP+L+KE R  
40 Sbjct: 121 QTLFNPNGSVLPQPRGPINEKLYARVELTDTQIKVDYFTDRHKLPSLSKEFPR 173

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1541

45 A DNA sequence (GBSx1632) was identified in *S.agalactiae* <SEQ ID 4749> which encodes the amino acid sequence <SEQ ID 4750>. This protein is predicted to be HAM1 family protein. Analysis of this protein sequence reveals the following:

Possible site: 23  
50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1218 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14796 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]

-1710-

Identities = 96/189 (50%), Positives = 130/189 (67%), Gaps = 1/189 (0%)

Query: 128 LIATHNKGRTKRFELPGKILGKVRNLNDYDPLFEVEETGMTTFEENARKLAEITSLKLTGK 187  
 +IATIN GK KEF+E+ G V++L + E+EEGT TFEENA +KAE ++K K  
 5 Sbjct: 8 LIATHNPGKVRKRFKLEPRGVDKSLAKIGTFEETIETGHTTFEENAIMKAEVAIAVANK 67

Query: 188 MVIDDSGLKVDALGGLPGVWSARFSGPDATTARNNAKLHRLAMVDFKERRSAQHITTL 247  
 MVI+DDSG +D LGG PGV+SAR++G D N K+L EL + +KE+R+A+F L  
 10 Sbjct: 68 MVIDADSGSLIDNLGGRPGVVSARYAGBQRDDQANTKVLSEIKGI -EKQRTARFCAL 126

Query: 248 VVSAPNKRSLVVEARWPGYIGTEPKGENGFYDPLFVVBGSRTAAELSAQKNLNSHRG 307  
 VS P +E+ VE GYI EP+GE GFGYDP+PIV + +T AEL++ EKN +SHR  
 15 Sbjct: 127 AVSIPGSRITKTVGHVHGVIABEPGRGYGFGYDPIFIVKDKKTMAEITSDKKNKISHRA 186

Query: 308 QAVRKIMRV 316  
 A++KL ++  
 20 Sbjct: 187 DALKKLSKL 195

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4751> which encodes the amino acid  
 20 sequence <SEQ ID 4752>. Analysis of this protein sequence reveals the following:

Possible site: 39

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2590 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 214/325 (65%), Positives = 253/325 (77%), Gaps = 5/325 (1%)

Query: 1 MTKTIFESKTEGWNFLGSFQAFNYFTCPG-NDESYEAIQGVFHRLLSTLKVE---GLQLH 56  
 M++ I+E K E NWF+G N + +G + + I D+ + +TL E G +  
 35 Sbjct: 14 MSEKIYKYDENWFIKRWTHNLISGNGVKHTTIKKIDLLDGLAATLWENPKGYIUS 73

Query: 57 VVQMISDFQLLAFVDMINGEYSRHIKVTQHGAILVSEDDQLFLVHLPKGTSLEKFPD 116  
 VV+ S I+ F+DMINGE R IRVT H G II+ E+ +L V+LP+ G S FF  
 Sbjct: 74 VVRHQSPLSLITFIIDMINGETQREIKVTFHAGTILMENAKLLAVLPLGGVSTATFF- 132

Query: 117 LKQNNPQDITILLIATNBEKTKFRELPGKILGKVRNLNDYDPLFEVEETGMTTFEENARL 176  
 ++ PGD IILAT NBEKTKFELPGKILGKVRNLNDYDPLFEVEETGMTTFEENARL  
 40 Sbjct: 133 ATSEQGFQDITILLIATNBEKTKFRELPGKILGKVRNLNDYDPLFEVEETGMTTFEENARL 192

Query: 177 KAETISLTKGMVLSDDSGLKVDALGGLPGVWSARFSGPDATTARNNAKLHRLAMVDFK 236  
 KAETIS-LTKGMV++DDSGLKVDALGGLPGVWSARFSGPDATTARNNAKLHRLAMVDFK  
 45 Sbjct: 193 KAETISLTKGMVLADDSGLKVDALGGLPGVWSARFSGPDATTARNNAKLHRLAMVDFK 252

Query: 237 BRSSAQPHITTLVVSAPNKRSLVVEARWPGYIGTEPKGENGFYDPLFVVBGSRTAAEL 296  
 ++BSAQPHITTLV+APNK+SLVVEA+WPGYI TP-KGENGFYDPLFVVBGSRTAAEL  
 50 Sbjct: 253 KKREAQPHITTLVVAAPNKRSLVVEA+WPGYIATQPKGENGFYDPLFVVBGSRTAAEL 312

Query: 297 AQEKNNLSHRGQAVRKIMRVFPKQ 321  
 A +KN LSHRGQAVRKIMRVFPKQ  
 55 Sbjct: 313 ADQKNNLSHRGQAVRKIMRVFPKQ 337

Based on this analysis, it was predicted that these proteins could be useful antigens for  
 vaccines or diagnostics.



-1711-

## Example 1542

A DNA sequence (GBSx1633) was identified in *S.agalactiae* <SEQ ID 4753> which encodes the amino acid sequence <SEQ ID 4754>. This protein is predicted to be glutamate racemase (murl). Analysis of this protein sequence reveals the following:

```

5      Possible site: 45
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.86    Transmembrane    114 - 130 ( 114 - 130)

10     ----- Final Results -----
      bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10067> which encodes amino acid sequence <SEQ ID 10068> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
      Identities = 160/267 (59%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

20     Query: 27  MDSRFIFLDSGVGGLTVVKEMFRQLPFESEVIFIGDQARAPYGFPRFAQQRIFBTFMWNVF 86
      M + IGF+DSGVGGLTVVKE RQLP E + ++GD AR FYGFRP Q+R+FTW+M +F
      Sbjct: 1  MKKQALGFIDSQVGLTVVKEAMRQLPNEISYYVGDTRCPYGFPRFEDQVRKFTWETHF 60

25     Query: 87  LLTKNVMIVFACNTATAVANVEVKAALDIPVLGVLPAGASAAIKSTNLGRKVGILGTPT 146
      LL RN+RN+VIACNTATA R ++IK+KL IPV+GVILPG+ AAIK+T+ ++G+IGT T
      Sbjct: 61  LLDENIKMLVFIACNTATAALDKDKKLAIPVIGVILPGSRAAIKATHTNIGVIGTGT 120

30     Query: 147  VESDAYRQKICALSPNTAVVSLACPFPVPIVSSNQSSSLAKKVVYETLSPLNGK-LDTL 205
      VKS+ Y+ I + V SLACPFPV+VSSN+ SS+AKKVY ETL EL + LDTL
      Sbjct: 121  VESNQTCKMHSKDTKALVTSLACPFPVFLVSSNEYSSAIAKKVVAETLRLNLEGLDITL 180

35     Query: 206  ILGCTHYFLLRPIIQNVMAEVLKIDSGAETVRDISVLLNYFRIDRWQNHQ-GRHHFYT 264
      ILGCTHYFLLRPIIQ +G V LIDSGAETV ++S +L+YF + + QNK +FYT
      Sbjct: 181  ILGCTHYFLLRPIIQNTLSDSVTLIDSGAETVSEVTILDYFNLAVDSSQVKEARNFYT 240

      Query: 265  TASPFGPKETAEQNLIS-QEINVERIVL 290
      T S + F I A +NL ++ VE I L
      Sbjct: 241  TGSSCMFHAIASENLQLDDLAVERITL 267

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4755> which encodes the amino acid sequence <SEQ ID 4756>. Analysis of this protein sequence reveals the following:

```

      Possible site: 19
      >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -1.70    Transmembrane    88 - 104 ( 86 - 104)

45     ----- Final Results -----
      bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50

```

The protein has homology with the following sequences in the databases:

```

      >GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
      Identities = 149/267 (55%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

55     Query: 1  MDTRFIFGLDSGVGGLTVVCELIROLPHKEIKIVIGDSARAPYGFPRPKQIKETWELNWF 60
      M + IGF+DSGVGGLTVV E +RQLP+E I Y+GD+AR FYGFRP+ Q++FTW+E +F
      Sbjct: 1  MKKQALGFIDSQVGLTVVKEAMRQLPNEISYYVGDTRCPYGFPRFEDQVRKFTWETHF 60

      Query: 61  LLTKNVMIVFACNTATAVANVEVKAALDIPVLGVLPAGASAAIKSTTKGQVIGTPT 120

```

-1712-

```

      LL +N+RM+V ACNTATA A +++ L IPV+GV+LPG+ AAIK+T ++GVIGT T
Sbjct: 61 LLDKNIMQIACNTATAAALDKIKKLAIPVIGVILPGSRAAIKATHYNRIGVIGTSGT 120

5  Query: 121 VASDIYRKIKQLLAPSIQVRSIACPKFVPVIVESNEMCSSIAKKIVYDLSLAPLWGK-IDTL 179
      V S+ Y+K I V SLACPKFVP+VESNE S+IAKK+V ++L PL + +DTL
Sbjct: 121 VKSNQYKRMVHSDKTDALVTSIACPKFVPVIVESNEYSSATAKKVVAETLRPLKNEGLDTL 180

      Query: 180 VLGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLINYPDIN-GNYHQKAVEHRFFTT 238
      +LGCTHYPLLRPIIQN +G SV LIDSGAE V ++S +L+YF++ + ++ E F+T
10  Sbjct: 181 ILGCTHYPLLRPIIQNTLIDSVTLIDSGAETVSEVSTILDYFNLAVDQNKEKASRNFTT 240

      Query: 239 TANPSIFQBIASINWLK-QKINVEHVTL 264
      T + ++P IAS WL+ + VEH+TL
Sbjct: 241 TGSSQMFHAIASENLQLDDLAVERHITL 267

15

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 195/264 (73%), Positives = 231/264 (86%)

20  Query: 27 MDSRFIGFLDSGVGLTVVKEMFRQLPEEVEIFIGDQARAPYGPRPAQQIREFTWQWVNF 86
      MD+RPIGFLDSGVGLTVV E+ RQLP E+++IGD ARAPYGRP +QI+E+TW+VNF
Sbjct: 1 MDTRFIGFLDSGVGLTVVCELRLQPKHEKIVYIGDSARAPYGRPRPKQIKEYTWELVNF 60

      Query: 87 LLTKNQNMIVACNTATAVAMQEIKEKLDIPVLGVILPGASAAIKSTNLGKVGIICTPMT 146
      LL+ +NMVIMV ACNTATAVAM++E+K LDIPVLGV+LPGASAAIKST G+VG+IGTPT
25  Sbjct: 61 LLTKNQNMIVFACNTATAVAMVEVKAALDIPVLGVILPGASAAIKSTTKGQVIGTPTMT 120

      Query: 147 VKSDAYRQIKQIALSPNTAVVSLACPKFVPVIVESNEMCSSIAKKIVYETLSPLVGLKDLTL 206
      V SD YR+KIQ L+P+ V SLACPKFVPVIVESN+M SS+AKK+VY++L+PLVGK+DTL+
30  Sbjct: 121 VASDIYRKIKQLLAPSIQVRSIACPKFVPVIVESNEMCSSIAKKIVYETLSLAPLVGKIDTLV 180

      Query: 207 LGCTHYPLLRPIIQNVMGAEVKLIDSGAETVRDISVLINYPFINHNMWNKHGGHHFTTA 266
      LGCTHYPLLRPIIQNVMG VKLIDSGAE VRDISVLINYP+IN N+ K H P+TTA
Sbjct: 181 LGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLINYPDINGNYHQKAVEHRFFTTA 240

35  Query: 267 SPKGFKEIASQWLSQSEINVERIVL 290
      +P+ F+EIA WL Q+INVE + L
Sbjct: 241 NPEIFQBIASINWLKQKINVEHVTL 264

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1543

A DNA sequence (GBSx1634) was identified in *S. agalactiae* <SEQ ID 4757> which encodes the amino acid sequence <SEQ ID 4758>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 21
      >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL Likelihood = -11.36 Transmembrane 3 - 19 ( 1 - 27)

      ----- Final Results -----
50  bacterial membrane --- Certainty=0.5943 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

60  >GP:CA13675 GB:Z99113 alternate gene name: yoxG [Bacillus subtilis]
      Identities = 26/72 (36%), Positives = 42/72 (58%)

      Query: 1 MSITIIWILLIIVALGGLVGGFIARQIKRKGIEHPRIATPDAIRNMSQMGKPSRAKV 60
      M++ + IL+ +VAL G+ G FIARK + + ++P + +R NM QMG KDS+ K+
Sbjct: 1 MTLNVGILGVGVALLIGVALGFFIARKYMSYLLKKNPPINEQMLRNMMNMGMKPSQKKI 60

      Query: 61 CQTYRNIVKHAK 72

```

-1713-

Q + + K  
 Sbjct: 61 NQMMKAMNNQTK 72

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4759> which encodes the amino acid sequence <SEQ ID 4760>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -10.72 Transmembrane 7 - 23 ( 1 - 27)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5288 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 62/79 (78%), Positives = 69/79 (86%)

Query: 1 NSITIWILLIIVALFGLVGGIFIAKQIEKEIGEHPRITPDAIREMMSQMGQKPSKAKV 60  
 MS INILL+IVAL G+ GGIFIAKQIEKEIGEHPRITP+AIKREMMSQMGQKPSKAK+  
 20 Sbjct: 1 MSTAINILLIIVALFGLVGGIFIAKQIEKEIGEHPRITPEAIREMMSQMGQKPSKAKI 60  
 Query: 61 QQTYRNIVKHAKTAIKTKK 79  
 QQTYRNII+K +K A+ K  
 25 Sbjct: 61 QQTYRNIIKQSKAAVSKGK 79

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1544

A DNA sequence (GBSx1635) was identified in *S.agalactiae* <SEQ ID 4761> which encodes the amino acid sequence <SEQ ID 4762>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.86 Transmembrane 82 - 98 ( 79 - 103)  
 35 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4142 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1545

45 A DNA sequence (GBSx1636) was identified in *S.agalactiae* <SEQ ID 4763> which encodes the amino acid sequence <SEQ ID 4764>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence  
 50 INTEGRAL Likelihood = -10.83 Transmembrane 56 - 72 ( 50 - 105)  
 INTEGRAL Likelihood = -7.27 Transmembrane 27 - 43 ( 17 - 48)  
 INTEGRAL Likelihood = -6.26 Transmembrane 76 - 92 ( 73 - 105)  
 INTEGRAL Likelihood = -4.83 Transmembrane 119 - 135 ( 118 - 141)  
 INTEGRAL Likelihood = -1.65 Transmembrane 160 - 176 ( 160 - 176)

-1714-

## ----- Final Results -----

bacterial membrane --- Certainty=0.5331(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8837> which encodes amino acid sequence <SEQ ID 8838> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4765> which encodes the amino acid sequence <SEQ ID 4766>. Analysis of this protein sequence reveals the following:

## Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

15 INTEGRAL Likelihood = -10.99 Transmembrane 45 - 61 ( 37 - 94)  
 INTEGRAL Likelihood = -7.06 Transmembrane 74 - 90 ( 62 - 94)  
 INTEGRAL Likelihood = -3.45 Transmembrane 110 - 126 ( 108 - 130)  
 INTEGRAL Likelihood = -2.18 Transmembrane 149 - 165 ( 149 - 165)  
 INTEGRAL Likelihood = -1.91 Transmembrane 21 - 37 ( 20 - 37)

## 20 ----- Final Results -----

bacterial membrane --- Certainty=0.5394(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/173 (64%), Positives = 145/173 (83%)

30 Query: 12 MSKKTTONVSTSLVAFAINIPIMPAKIIIGPASFTLASHVPLFLSIFISVFVAIIA 71  
 M+KK TQ+++YTSILVAFAI+PIIMP K+IIGPASFTLASHVPLFL+IP+S+PVAIIA  
 Sbjct: 1 MTKKFTQLIATYSILVAFALPIIIMPLKLIIGPASFTLASHVPLFLAIFMSIPVAIIA 60

Query: 72 LGTGLGFLLAGFPFIVIVLRALSHIGLPALIAAPLIKSKPILMSKWKQTLFLAVAINIHGL 131  
 LGT LGFLLAG P++IVLRALSH+ FA++AA+ + KP L+ S + FA IN+IHGL  
 35 Sbjct: 61 LGTTLGFLLAGPLITIVLRALSHLLFAIIAAMWLGRKQLMTSAVKCFSPAFFINIVHGL 120

Query: 132 LEFTIVYIITMTSNSSTYIMSLPFLIGLGLSHGLVDFYIALFIWKKTKQL 184  
 EF+ VYI+T T+ +S +Y WS+ LIGLGL+HG++DFY+AL +N+++ + L  
 40 Sbjct: 121 AEFLVYIILTATATSMGYFWMGLGILGLGLSHGLVDFYIALVLRFLAKNL 173

A related GBS gene <SEQ ID 10789> and protein <SEQ ID 10790> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 3

SRCLIG: 0

45 MCG: Length of UR: 24  
 Peak Value of UR: 3.16  
 Net Charge of UR: 2  
 MCG: Discrim Score: 12.56  
 GVR: Signal Score (-7.5): -0.16

## 50 Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

Amino Acid Composition: calculated from 20

ALOM program count: 5 value: -10.83 threshold: 0.0

55 INTEGRAL Likelihood = -10.83 Transmembrane 45 - 61 ( 39 - 94)  
 INTEGRAL Likelihood = -6.26 Transmembrane 65 - 81 ( 62 - 94)  
 INTEGRAL Likelihood = -4.83 Transmembrane 108 - 124 ( 107 - 130)  
 INTEGRAL Likelihood = -1.65 Transmembrane 149 - 165 ( 149 - 165)  
 INTEGRAL Likelihood = -0.27 Transmembrane 24 - 40 ( 24 - 40)  
 PERIPHERAL Likelihood = 0.42 86

-1715-

modified ALOM score: 2.67  
icml HYPID: 7 CFP: 0.533

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5331(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1546

A DNA sequence (GBSx1637) was identified in *S.agalactiae* <SEQ ID 4767> which encodes the amino acid sequence <SEQ ID 4768>. This protein is predicted to be transcriptional regulator, biotin repressor family. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2237(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14749 GB:Z99118 ytxA [Bacillus subtilis]

Identities = 72/165 (43%), Positives = 112/165 (67%), Gaps = 2/165 (1%)

Query: 6 RRENILITLKGTEKRAISASTLAKIPSVSRQVIVGDIALLRQQCDIISTPKGYL-MSSAL 64  
RR++L LK +K ++ LAK +VSRQVIV DI+LLA++ I1+T +GI+ M +A  
Sbjct: 12 RRDQLLLMLKRSKSPITGGELAKKANVSQVIVQDISLKKANVPIIATISQVYVMDAAA 71

Query: 65 STHOFTARLV-COHGIEQTEBELIILRYQGIIMVVEVHPYGMLTAPLNISQKQKIDIN 123  
HQ R++ C HS E+TEBEL--I+ + +V++RHP+YG LIA + + ++K++ +  
Sbjct: 72 QHQQAERIIACLGHPERTTEBELQLVDEGVTVDKVKIHPVYGDLTAAIQVGRKEVSH 131

Query: 124 FTAKIKVSNALLSSLTGXLHTHMISQDQSVDFQICEALKKAGI 168  
F K+ +NA LS LTG+H H ++ D+ DQ C+AL++AGI  
Sbjct: 132 FTKINSTNAAYLSQTDGVLHLLTAPFDEHRIDQACQALEEAGI 176

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4769> which encodes the amino acid sequence <SEQ ID 4770>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2971(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 109/170 (64%), Positives = 136/170 (79%)

Query: 1 MKAQERRENILITLKGTEKRAISASTLAKIPSVSRQVIVGDIALLRQQCDIISTPKGYIM 60  
MKA++RR+ I+ L ++A+SA+ L K+ VSRQVIVGDIALLRQQ DIISTPKGY+M  
Sbjct: 1 MKAEDRRQKIIECLNSBOKAVSATRLGKLLGVSQVIVGDIALLRQQDIISTPKGYIM 60

Query: 61 SSALSTHGTARLVCOHGIETEELEIILRYQGIIMVVEVHPYGMLTAPLNISQKQD 120

-1716-

S+AL +HQP AR+VCOH +E+T++ELEILL +QGII VEVSHPIYGM+TAPLNI++ D  
 Sbjct: 61 STALYSHQPCARIVCOHNVETKKEILLAHQGIITTVVEVSHPIYGM+TAPLNIKTHSD 120

Query: 121 IDNFTAKLVSNARLLSSLTGLATHMISCDQSVFQICRALKAGILY 170  
 + NF +KL S AKLLSSLT+GLH+H+ISC Q P I L+ AGIIL  
 Sbjct: 121 VTNFMKLSQSKARLLSSLTGLASHLISCPQEAFLAKHDLKAGIILY 170

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 10 Example 1547

A DNA sequence (GBSx1638) was identified in *S. galactiae* <SEQ ID 4771> which encodes the amino acid sequence <SEQ ID 4772>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

15 INTEGRAL Likelihood = -8.44 Transmembrane 143 - 159 ( 138 - 165)  
 INTEGRAL Likelihood = -8.17 Transmembrane 164 - 180 ( 160 - 184)  
 INTEGRAL Likelihood = -7.17 Transmembrane 56 - 72 ( 53 - 78)  
 INTEGRAL Likelihood = -5.63 Transmembrane 24 - 40 ( 21 - 44)  
 INTEGRAL Likelihood = -4.94 Transmembrane 113 - 129 ( 108 - 131)

20 INTEGRAL Likelihood = -2.39 Transmembrane 86 - 102 ( 86 - 103)  
 INTEGRAL Likelihood = -1.06 Transmembrane 203 - 219 ( 203 - 219)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4376(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10069> which encodes amino acid sequence <SEQ ID 10070> was also identified.

### 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC18360 GB:AF064763 putative membrane spanning protein  
 [Lactococcus lactis subsp. cremoris]  
 Identities = 97/188 (51%), Positives = 133/188 (70%)

35 Query: 38 IMLYMFPONMIAIMKMGILYFGAILLELVLFVFSGAARRNTPAALPLFLIYSALNGFT 97  
 IM+ F NM AL+Q I+ LV+V GA +N+ ALP+F+ Y+A GF  
 Sbjct: 1 IMITFFLDNMRAILQSGSLFLVLVLIPLVMVVSQGLAMKNSKALPFIYGAAFMGFL 60

40 Query: 98 LSFIDARYTOTTVLOAFITSAAVFFAMALIGAKTKIDLSGMRKALMAALIGILIASLVNL 157  
 +SF + YT T + APIT++A+FF +++ G TK+LSGM KAL A+ G++A L+NL  
 Sbjct: 61 ISFTLLMYTATDITLAFITASAMF+GLSVYGRFTKRNLSGMRKALGVAWGLIVAMNLNL 120

Query: 158 FIGSGMSYIISIVCVII FSGLIAYDNOMIKYVNSQGGVADGWAVSMALSLYLDFINL 217  
 F S G++ +IS+V V+IFSGLIA+DNQ I VYN+ GQV+DGWA+SMALSLYLDFINL  
 45 Sbjct: 121 FFASGTGLTILISLVGVVIFSGLIANDNOKITQVYNHNGQVSDGWAISMALSLYLDFINM 180

Query: 218 FLNIRLF 225  
 FL +LRLF  
 Sbjct: 181 FLFLIRLF 188

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4773> which encodes the amino acid sequence <SEQ ID 4774>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

55 INTEGRAL Likelihood = -8.97 Transmembrane 143 - 159 ( 138 - 165)  
 INTEGRAL Likelihood = -5.89 Transmembrane 164 - 180 ( 160 - 184)  
 INTEGRAL Likelihood = -5.68 Transmembrane 56 - 72 ( 55 - 77)  
 INTEGRAL Likelihood = -4.78 Transmembrane 113 - 129 ( 110 - 130)

-1717-

INTEGRAL	Likelihood = -2.81	Transmembrane	203 - 219 ( 203 - 222)
INTEGRAL	Likelihood = -2.76	Transmembrane	24 - 40 ( 23 - 41)
INTEGRAL	Likelihood = -2.75	Transmembrane	86 - 102 ( 86 - 104)

5 ----- Final Results -----  
           bacterial membrane --- Certainty=0.4588 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:AA018360 GB:AF064763 putative membrane spanning protein  
           [lactococcus lactis subsp. cremoris]  
           Identities = 90/189 (47%), Positives = 133/189 (69%)

15 Query: 38 INLYPFRENILISILVNPQMIYYGAAITELILVFPVASSAARKNTPAALPIFLIYSALNGFT 97  
           +N+ F +N+ +IL + + II L++V A KN+ ALPIF+ Y+A GF  
           Sbjct: 1 IMITTFPLDNRAILQSGSLFLVLWIIPLVMVVSQGLAMKNSKMALEPIFGYAAFMGFL 60

20 Query: 98 LSPFIIVAYQTTVFQAFSSAAVFPMSIIGVTKRMSGLRKAMPAALIGVVVASLINL 157  
           +SF ++ Y T + AF++++A+FF +S+ G TKR++SG+ KA+ A+ G++VA L+NL  
           Sbjct: 61 ISPTILMTTATDITLITAFASAMFFGLSVYGRFTRKNLSGNGKALGVAVNGLIVAMLLNL 120

Query: 158 FIGSGMMSYVISVISVLIPSGLIASDNQMIKRVYQATNGQVDGWAVAMALSLYLDFINL 217  
           F S ++ +IS++ V+IPSGLIA DNQ I +VY A NGQV DQWA++MALSLYLDFIN+  
 25 Sbjct: 121 FFASTGLTILISLVGVIPSGLIANDNQKLTQVYNHNGQVSDGWAISMALSLYLDFINM 180

Query: 218 FISLIRIFG 226  
           F+ LIR+FG  
           Sbjct: 181 FLFLRLRFG 189

30

An alignment of the GAS and GBS proteins is shown below.

Identities = 167/229 (72%), Positives = 202/229 (87%)

35 Query: 1 MNDNVITYQSDGSLNQFFAKIYGLVGIGVGLSAAVSAIMLYPFPQNMIAMQMPGLYFS 60  
           MND+VITYQSD GLNQFFAKIY LVG+GVGLSA VS +MLY F +N+I+I+ P +Y+G  
           Sbjct: 1 MNDNVITYQSDVGLNQFFAKIYSLVGMGVGLSAFVSYIMLYPFPRENILISILVNPQMIYYG 60

Query: 61 AITELILVFPVASSAARKNTPAALPLFLIYSALNGFTLSPIIARYQTTLVQLAFITSAV 120  
           A I+EL+LVFPVAS AAR+NTPAALP+FLIYSALNGFTLSPII Y QTTV QAF++SAV  
 40 Sbjct: 61 AAITELILVFPVASSAARKNTPAALPIFLIYSALNGFTLSPIIIVAYQTTVFQAFISSAAV 120

Query: 121 FFAMALIGAKTKKDLGMRKALMAALIGILIASLVNLFPGSGGMSYIISIVCVIIPSGLI 180  
           FFAM++IG KTK+D+SG+RKA+ AALIG+++ASL+NLPIGSG MSY+IS++ V+IPSGLI  
 45 Sbjct: 121 FFAMSIIGVTKRMSGLRKAMPAALIGVVVASLINLPIGSGMMSYVISVISVLIPSGLI 180

Query: 181 AYNQMIKYVYNSQGGQVADGWAVSMALSLYLDFINFLINILRFPARN 229  
           A DNQMIK VY + QGV DQWAV+MALSLYLDFINLFL++LR+P RND  
           Sbjct: 181 ASDNQMIKRVYQATNGQVDGWAVAMALSLYLDFINLPIILLRFPORN 229

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1548

A DNA sequence (GBSx1639) was identified in *S. agalactiae* <SEQ ID 4775> which encodes the amino acid sequence <SEQ ID 4776>. Analysis of this protein sequence reveals the following:

55 Possible site: 44  
       >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2495 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

60

-1718-

A related GBS nucleic acid sequence <SEQ ID 10071> which encodes amino acid sequence <SEQ ID 10072> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4777> which encodes the amino acid sequence <SEQ ID 4778>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3277 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 127/163 (77%), Positives = 141/163 (85%)

- Query: 7 YQDDKDFNDLVGHLIDHPRFQKLEAIVQRHHSTRLEHSINVSYTSYKIAKKFGWDASSTA 66  
Y +DK++N+ VGHLI HRFQKL IVQH HSTLEHSINVSY+SYK+AK+FGWDA STA  
20 Sbjct: 3 YTEDKRYMEHVGHILAHFRFQKLSHIVQHCHSTRLEHSINVSYSSYKIAKKFGWDAKSTA 62
- Query: 67 RGGLLEDFYYDNRVTKFNKSHAWHPRIAVRNARKLTDLNAREEDILKHMWGATIAFP 126  
RGGLLEDFYYDNRVTKFNK HAWHPRIAVRNA+KLT+LN +EEDILKHMWGATIA F  
25 Sbjct: 63 RGGLLEDFYYDNRVTKFNKSHAWHPRIAVRNARKLTDLNAREEDILKHMWGATIAFP 122
- Query: 127 RYKESYIVTWDKYNAVREASRFLKRIFFKKPIRFSRFLGSHN 169  
RYKESYIVTWDKYNAV+EA PL++ + RK L SHN  
Sbjct: 123 RYKESYIVTWDKYNAVRAVTPLRQGSNRRFLRKLTGSHN 165

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1549

A DNA sequence (GBSx1640) was identified in *S.agalactiae* <SEQ ID 4779> which encodes the amino acid sequence <SEQ ID 4780>. Analysis of this protein sequence reveals the following:

- 35 Possible site: 37  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -3.03 Transmembrane 213 - 229 ( 212 - 229)
- 40 ----- Final Results -----  
bacterial membrane --- Certainty=0.2211 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9413> which encodes amino acid sequence <SEQ ID 9414> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CB14825 GB:Z99118 similar to rRNA methylase [Bacillus subtilis]  
Identities = 96/228 (42%), Positives = 143/228 (62%), Gaps = 5/228 (2%)

- 50 Query: 3 QKRYKSSYLIGWHLPEEARCKYGAQPLNIFVT-ETAI DR-LRKPERALVITVDVLKELT 60  
+++ + +++LIEG HL EEA K I V ET I L + ++++D +T  
Sbjct: 22 KERTKTNTPLTEGSHLVSEALKSPGIKVLVDETRIPSDLETIGICTMISEDAPSAVT 81
- Query: 61 DSQTPQGITVAEIAFQETRWTDIKKGRFLVLEVDQDFGNLGTWVKTADAANFADVFLSQXS 120



-1719-

+++TPQ I A E + +K L+++ VQDFGNLGT++RTADAA DAV L +  
 Sbjct: 82 ETETPQQIAAVCHMPFEEKLATARK--VLLIDAVQDFGNLGTMI RTADAGLDAVLGSGT 139  
 Query: 121 ADLYNQKTLRSMQGSHPLFVFRVEIQFVFNCKABGITMIATTLSEQSVNYKLNPKVDY 180  
 AD +N KTLRS QGSHPH+PV R + +V+ KAEG+ T L + Y+ +P+ +  
 Sbjct: 140 ADAPNGKTLRSQGSHPHIIPVVRNRLPSTVDLKAEGVKVYGTAL-QNGAPYQEI PQSES 198  
 Query: 181 PALIMGNEGQGISKXTMTBEADVLAHIEMPGQASINAVAAAGVVFSL 228  
 PALI+GNEG G+ + E+ D+ ++ + GQASINAVAA +++++ L  
 Sbjct: 199 FALIVNEGAGVDAALLETLDNLNVLVPLTGQASINAVAAAILVYHL 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4781> which encodes the amino acid  
 sequence <SEQ ID 4782>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.97 Transmembrane 229 - 245 ( 228 - 245)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2190(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 141/229 (61%), Positives = 178/229 (77%)  
 Query: 1 MLQKKYRKSSYLIEGWHLFERAKYGAQFLNIFVITADRLRKPRAIVTDDVKELT 60  
 +LQKK+RK SYLIEGWHLFEEA+K G F +IFV E ++RL + ++V+ VLKELT  
 Sbjct: 17 LQKKHRRQSYLIEGWHLFEEAQQSGQVFRHIFVLEENVERLAGEQELVIVSPQVLKELT 76  
 Query: 61 DSQTFQQIVARIAFQETRWTDINKRFLVLEDVQDFGNLGTMTADAAFDVFLSCKS 120  
 DS +FQIVAB+ + + KG++LVLEDVQDFGNLGT++RTADAA FD VFSL+KS  
 Sbjct: 77 DSSPQQIVABVEIPIKLAFFSDYKGYLVLEDVQDFGNLGTII RTADAAFDGVLSEKS 136  
 Query: 121 ADLYNQKTLRSMQGSHPLFVFRVEIQFVFNCKABGITMIATTLSEQSVNYKLNPKVDY 180  
 AD+YNQKTLRSMQGSHPLF++R ++ Q + ++ATTLG++SV+YK+L ++  
 Sbjct: 137 ADLYNQKTLRSMQGSHPLFIWRTDVTQLCRLQGYETPILATTLSEKSVDYKELTHHR 196  
 Query: 181 PALIMGNEGQGISKXTMTBEADVLAHIEMPGQASINAVAAAGVVFSLI 229  
 AL++GNEGQGIS N AD L HT MPQASINAVAAAG++IFSLI  
 Sbjct: 197 LALVLNEGQGISAEHALADQLVHTMPQASINAVAAAGILIFSLI 245

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

A related GBS gene <SEQ ID 8839> and protein <SEQ ID 8840> were also identified. Analysis of this  
 protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7  
 McG: Discrim Score: -7.98  
 GvH: Signal Score (-7.5): -3.86  
 Possible site: 37  
 >>> Seems to have no N-terminal signal sequence  
 ALON program count: 1 value: -3.03 threshold: 0.0  
 INTEGRAL Likelihood = -3.03 Transmembrane 213 - 229 ( 212 - 229)  
 PERIPHERAL Likelihood = 5.14 149  
 modified ALON score: 1.11  
 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1720-

The protein has homology with the following sequences in the databases:

[illegible]

35 SEQ ID 8840 (GBS430) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 5; MW 29kDa).

GBS430-GST was purified as shown in Figure 220, lane 8.

### Example 1550

A DNA sequence (GBSx1641) was identified in *S. agalactiae* <SEQ ID 4783> which encodes the amino acid sequence <SEQ ID 4784>. This protein is predicted to be acylphosphatase (acyP). Analysis of this protein sequence reveals the following:

Possible site: 48  
>>> Seems to have an uncleavable N-term signal seq

```

45  ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < success
      bacterial outside --- Certainty=0.0000 (Not Clear) < success
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < success

```

50 - A related GBS nucleic acid sequence <SEQ ID 10073> which encodes amino acid sequence <SEQ ID 10074> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD36630 GB:AE001801 acylphosphatase, putative [Thermotoga maritima]  
Identities = 35/88 (39%), Positives = 52/88 (58%), Gaps = 3/88 (3%)

Query: 24 MKKVLHIVSGRVQGVGFYATYSLALFEGIDYGRVWNNDDGTVEILAQSTDSNKMVTFIQ 83  
MK + + V G VQGVGFY T +A +G + G V N DDG+V I A+ D N + +E+

-1721-

Sbjct: 1 MKALKIRVEGIVQGVGFRYPTRRVAKSLG-VKGYVMNDGVSFIHAEG-DENALRRFLN 58

Query: 84 KIRKGPSKWSKVTVYDIKLNFDFNDF 111

++ KGP + VT V ++ + + DF

Sbjct: 59 EVAGPPFA-AVVINVSVRETTPEGYEDF 85

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4785> which encodes the amino acid sequence <SEQ ID 4786>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 69/95 (72%), Positives = 85/95 (88%)

Query: 19 KRGCVMKCVHLIVSGRVQGVGFRYATYSLEIGDIYGRVWNNDDGTVEILAQSTDENM 78

K +\*+K\*V LIVSGRVQGVGFRYAT++LAL+IGDIYGRVWNN+DGTVEILAQS D\*+K+

Sbjct: 7 KEALLMCKVRLIVSGRVQGVGFRYATHLTALDIDIGYGRVWNNDDGTVEILAQSKDEDKI 66

Query: 79 TOPFIQKIRKGPSKWSKVTVYDIKLNFDFNDFPM 113

FIQ++RKGPSKWSKVTVYD+ + NF+DF DF++

Sbjct: 67 ATFIQSVRKGPSKWSKVTVYDVIMANFDFQDFQI 101

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1551

A DNA sequence (GBSx1642) was identified in *S.galactiae* <SEQ ID 4787> which encodes the amino acid sequence <SEQ ID 4788>. This protein is predicted to be membrane protein homolog (yidC). Analysis of this protein sequence reveals the following:

Possible site: 16

>>> May be a lipoprotein

INTEGRAL Likelihood = -12.52 Transmembrane 60 - 76 ( 54 - 83)

INTEGRAL Likelihood = -3.66 Transmembrane 178 - 194 ( 177 - 196)

INTEGRAL Likelihood = -2.76 Transmembrane 140 - 156 ( 137 - 157)

INTEGRAL Likelihood = -2.60 Transmembrane 216 - 232 ( 213 - 232)

----- Final Results -----

bacterial membrane --- Certainty=0.6010(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10075> which encodes amino acid sequence <SEQ ID 10076> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF03934 GB:AF139908 membrane protein homolog [Listeria

monocytogenes]

Identities = 82/222 (36%), Positives = 133/222 (58%), Gaps = 4/222 (1%)

Query: 44 PMANLITYFAQHGLGFGVALIIIVTVIVRVVILPLGLYQSWKASVQAKRMAYFKPLFEPI 103

P + I + A+ G +G+ALII T++R +I+PL L + KMA KP + I

Sbjct: 3 PPTSPIMFVAKFVGNGYIALIITTLIRALIMPLNLRRTAKAQMGWSKMAVAKPEIDBI 62

-1722-

Query: 104 NERLRNNAKTQEEKLAQOETELMTAQRENGSLSPGGIGCLPLLIQMPFFSAIFFAARYTPGV 163  
 RL+ A ++EE+ Q E+M + ++ +GCLPLLIQMP A ++A R + +  
 Sbjct: 63 QARLKRAATSEBQRTIQEEMNAVYSKYINPMQ-NGCLPLLIQMPILMAFYAIRGSSEI 121

5 Query: 164 SSATFLGLNLQKSLTLTVIIAALYFVQSWLSMOGVDPBQRQCKMTMYLMPIMNVFMSI 223  
 +S TFL NLG + L +I ++Y Q ++SM G EQ++QMK + +PIN++F+S  
 Sbjct: 122 ASHTFLMFLNLSGSDPMLVLAITAGLVLAQYFVSMIGYSPBQKQCKIIGLMSPTIMLVFS 181

10 Query: 224 SLPASVALYWFIGIFSIIOQLVT--TYVLK-PKLRKKVEHE 262  
 + P+++ALYW +GG+P Q L+T Y+ K P+++ +EE  
 Sbjct: 182 TAPGALALYWAGVGLFLAGQTLTKLYMKNHPEKVMQSE 223

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4789> which encodes the amino acid sequence <SEQ ID 4790>. Analysis of this protein sequence reveals the following:

15 Possible site: 31

>>> May be a lipoprotein

20 INTEGRAL Likelihood = -9.55 Transmembrane 62 - 78 ( 54 - 82)  
 INTEGRAL Likelihood = -2.81 Transmembrane 178 - 194 ( 177 - 195)  
 INTEGRAL Likelihood = -0.90 Transmembrane 216 - 232 ( 215 - 232)

----- Final Results -----

25 bacterial membrane --- Certainty=0.4621(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>QF:AAF03934 GB:AF139908 membrane protein homolog [Listeria monocytogenes]  
 Identities = 89/218 (40%), Positives = 132/218 (59%), Gaps = 2/218 (0%)

30 Query: 43 KPMSTYIDYFANNAGLYGGLAIIVTIVRTLLPLGLYQSWKASYQSEKMAFLKVFEP 102  
 +P + FI + A G YG+AIII T+++R LI+EL L + KMA KP +  
 Sbjct: 2 QPSTSPIMFAKPVGGNYGIAIITITLIRALIMELNETAKAQMGMSKMAVAKFEIDE 61

35 Query: 103 INKRIQANSQEEKMAAQTELMAAQRANGINPLSGIGCLPLLIQMPFFSAMFYAAQYTKZ 162  
 I R+KA S+EE+ Q E+M + INP+ +GCLPLLIQMP A Y+A + +  
 Sbjct: 62 IQARLKRAATSEBQRTIQEEMNAVYSKYINPMQ-NGCLPLLIQMPILMAFYAIRGSSE 120

40 Query: 163 VSTSTFMGIDLSRSILVLTATIAALYFQSWLSMOGVDPBQRQCKMTMYLMPIMNVFMS 222  
 +++ TP+ +LGS +VL I ++Y Q ++SM+ S EQ++QMK + PIM++F+S  
 Sbjct: 121 IASHTFLMFLNLSGSDPMLVLAITAGLVLAQYFVSMIGYSPBQKQCKIIGLMSPTIMLVFS 180

45 Query: 223 FSLPAGVGLYMLVGGFISIIQQLITTYLLKPLRHQIK 260  
 F+ P+ + LYW VGG F Q L+T L + H +IK  
 Sbjct: 181 TAPGALALYWAGVGLFLAGQTLTKLYMKNHPEK 217

An alignment of the GAS and GBS proteins is shown below.

Identities = 203/309 (65%), Positives = 254/309 (81%), Gaps = 2/309 (0%)

50 Query: 1 MKTKTKRILFSSILSLMLLLLTGCVSDGKAGKPYGVINWTLGVPMANLITYFAHQGLGF 60  
 +K TL RILFS L+L+LL LGVC D G P G+IW LG PM+ I YFA + GLG+  
 Sbjct: 1 LKLTINRILFSGILSLILLLTIGCVGRDAHGNKGMINEFLGKPMSTYIDYFANNAGLY 60

55 Query: 61 GLAIIIVTVIVRVVILPLGLYQSWKASYQSEKMAFKLVFEPINERLRNNAKTQEEKLAQ 120  
 G+AIIVT+IVR +ILPLGLYQSWKASYQ+EKD+ KP+FEPIIN+R++ A +QSEK+AAQ  
 Sbjct: 61 GLAIIIVTVIVRTLLPLGLYQSWKASYQSEKMAFKLVFEPINERLRNNAKTQEEKLAQ 120

60 Query: 121 TELMTAQRENGSLSNKRGICGLPLLIQMPFFSAIFFAARYTPGVSSATFLGLMLQKSLTL 180  
 TELM AQR +G++ GSGICGLPLLIQMPFFSAIFFA+YT GVS++T+G++LG +SL L  
 Sbjct: 121 TELMAAQRANGINPLSGIGCLPLLIQMPFFSAMFYAAQYTKGVSTSTFMGIDLSRSILV 180

Query: 181 TVIIAALYFQSWLSMOGVDPBQRQCKMTMYLMPIMNVFMSISLPASVALYWFIGIFSI 240  
 T IIA LYF QSWLSM V +BQR+QCKMTMY MPIM+MS SLEA V LYW +GG FS  
 Sbjct: 181 TATIAALYFQSWLSMOGVDPBQRQCKMTMYLMPIMNVFMSISLPAGVGLYMLVGGF 240

-1723-

Query: 241 IIQQLVTTYVLAKPLRRKVEREYTKNPPKAYKANKARKDVNSTKATESNQAIITSKKTN 300  
 IIQQLVTTY+LKP+L +++REY KNPPKAY++ ++RKDV T S ++N + K+N  
 Sbjct: 241 IIQQLVTTYLKLPLRLKQIKREYAKNPPKAYQSTSSRKDVTPSQNMEQAN--LPKKIKSN 298

Query: 301 RNAGKQKRR 309  
 RNAGKQ+R  
 Sbjct: 299 RNAGKQKRR 307

- 10 A related GBS gene <SEQ ID 8841> and protein <SEQ ID 8842> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 23 Crend: 6  
 MoG: Discrim Score: 8.74  
 GVH: Signal Score (-7.5): -1.47  
 Possible site: 16  
 >>> May be a lipoprotein  
 ALOM program count: 4 value: -12.52 threshold: 0.0  
 INTEGRAL Likelihood = -12.52 Transmembrane 60 - 76 ( 54 - 83)  
 INTEGRAL Likelihood = -3.66 Transmembrane 178 - 194 ( 177 - 196)  
 INTEGRAL Likelihood = -2.76 Transmembrane 140 - 156 ( 137 - 157)  
 INTEGRAL Likelihood = -2.60 Transmembrane 216 - 232 ( 213 - 232)  
 PERIPHERAL Likelihood = 0.74 235  
 modified ALOM score: 3.00

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.6010 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

37.9/63.7% over 193aa

Bacillus subtilis

EGAD|45886| hypothetical 30.7 kD lipoprotein in glng-ansr intergenic region precursor  
 Insert characterized  
 SP|P54544|YQJG\_BACSU HYPOTHETICAL 30.7 KDA LIPOPROTEIN IN GLNG-ANSR INTERGENIC REGION  
 PRECURSOR. Insert characterized  
 GP|1303958|dbj|BAAL2613.1||D84432 YqjG Insert characterized  
 GP|2634823|emb|CAB14320.1||Z99116 similar to lipoprotein SpoIIJ-like Insert  
 characterized  
 PIR|G69963|G69963 lipoprotein SpoIIJ-like homolog yqjG - Insert characterized

CRF02470(478 - 1038 of 1530)

EGAD|45886|BS2384(63 - 256 of 275) hypothetical 30.7 kD lipoprotein in glng-ansr intergenic  
 region precursor [Bacillus subtilis]GP|P54544|YQJG\_BACSU HYPOTHETICAL 30.7 KDA LIPOPROTEIN  
 IN GLNG-ANSR INTERGENIC REGION PRECURSOR.GP|1303958|dbj|BAAL2613.1||D84432 YqjG [Bacillus  
 subtilis]GP|2634823|emb|CAB14320.1||Z99116 similar to lipoprotein SpoIIJ-like [Bacillus  
 subtilis]PIR|G69963|G69963 lipoprotein SpoIIJ-like homolog yqjG - Bacillus subtilis  
 %Match = 13.0  
 %Identity = 37.9 %Similarity = 63.7  
 Matches = 72 Mismatches = 65 Conservative Sub.s = 49

252 282 312 342 372 402 432 462  
 PCGSIV\*FLAKK\*NR\*VY\*KLESELTLLKTLKRIILPSSISLSMLLLITGCVSVDKAGKPTGVKINTLGVPMAHLITYFAQ

MLRTYQKLLAMGIFILVLCSGNAAPFATNQVGGLENVGFPHDYLIEPFSALLKGVAG  
 10 20 30 40 50

492 522 552 582 612 642 672 702  
 HQQLGFGVAILIVTVIVRVILPLGLYQSWKASYQAEKMAFKPLPDPINERLRNAKTOEKKLAQTELMTAQRENGLSM  
 :|::|||::|||::|||::| | |||| | : : | : : | : : :  
 LFKHEYGLSGLITVIVRVILPLVNPQFKKRIIPQSKMAVKKPCVDSIQVLKKTQPEKQKRLQEMWMLYQSHINDP  
 70 80 90 100 110 120 130

-1724-

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50      **Example 1552**

A DNA sequence (GBSx1644) was identified in *S. agalactiae* <SEQ ID 4791> which encodes the amino acid sequence <SEQ ID 4792>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

```

55      Possible site: 48
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.98    Transmembrane    32 - 48 ( 23 - 53)
      INTEGRAL    Likelihood = -9.18    Transmembrane    195 - 211 ( 189 - 213)
      INTEGRAL    Likelihood = -8.70    Transmembrane    72 - 88 ( 62 - 93)

60      ----- Final Results -----
      bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1725-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter  
(permease) [Bacillus subtilis]  
Identities = 116/217 (53%), Positives = 168/217 (76%)

Query: 2 INWDATFNLEKAVKAPPSVIQGLPYTIGLSLVGFILGAIVGFFVAMKMSHPRLLRYLAN 61  
I W+ IFN +LA+++FP VI+G+ YT+ +S V G ++G F++L +MS LLR+ A  
Sbjct: 5 IQWEYIPNFKLAIESPPVYIKGIVTLLISPVSMFAGTVIGLFISLARMKIALLRWPAK 64

Query: 62 IHISLMRGIPMLVILFLIYFGLPFFIGIQLDVASTIVGPFMMSSAYISEIRAAVLAVDH 121  
++IS MRG+P++V+LF++YFG P+IGI+ AVTA+++GF++ S+AYI+EI R+A+ +V+  
Sbjct: 65 LYSIFMRGVFILVILFILYFGPFYIGFISAVTAALIGFSLNSAAYIAEINRSIAISSVEK 124

Query: 122 GQWEAARALGLKTPPIYRGIIPOATRIALPSLNVLLDMVKSSLTAMITVPDIFNNAK 161  
GQWEAA -LGL RGII+PQ+ RIALP L+NVLLD++K+SSL AMITVP++ +AK  
Sbjct: 125 GQWEAASSLGLSYQTMRGIIIPQSTRIALPPLANVLLDLIKASSLAAMITVPPELLQPAK 164

Query: 182 IVGGTYSYDMYAYILVALIYVICTLYAIQDWKEK 218  
I+GG DYMT YIL ALIYW IC++ A+ Q+ EK+  
Sbjct: 185 IIGGREFDYMTMYILTAIYWAICSIARVFNILEKK 221

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4793> which encodes the amino acid sequence <SEQ ID 4794>. Analysis of this protein sequence reveals the following:

Possible site: 23  
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood	Transmembrane	186 - 202 ( 184 - 205)
INTEGRAL	Likelihood = -5.79	Transmembrane	26 - 42 ( 21 - 43)
INTEGRAL	Likelihood = -5.84	Transmembrane	57 - 73 ( 56 - 84)
INTEGRAL	Likelihood = -4.78	Transmembrane	86 - 102 ( 86 - 103)

----- Final Results -----  
bacterial membrane --- Certainty=0.3718(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter  
(permease) [Bacillus subtilis]  
Identities = 113/214 (52%), Positives = 157/214 (72%)

Query: 1 MINIPMLKDSLGFVLSGLPYTIGISLSFPTGLFGLGLALLGRSROPLIHLYVRAYISI 60  
+ N L +S +V+ YTL IS +S P G +GL ++L S+ L+ + +YIS  
Sbjct: 10 IFNFKLAIESPPYIKGIVTLLISPVSMFAGTVIGLFISLARMKIALLRWPAKLYISF 69

Query: 61 MRGVPMIVLVFLVLYFGLPYGLEGFPALLCAYLGFMSVSAAYISEVFRSSIEADKQWEA 120  
MRGV++V+LF+LYFG PY G+E A+ A +GPS+ SAAYI+E+ RS-I +++KQWEA  
Sbjct: 70 MRGVPLVILVILYFGLPYGIFESAVTAALIGFSLNSAAYIAEINRSIAISSVEKQWEA 129

Query: 121 AVALGLPYALMKVKIILPQAFRIAPVPLGNVIMVKSSLSAAMITVPDIPQNAKIIGR 180  
A +GL Y ++ IILPQ+ RIA+PPL NV++D++K+SSLAAMITVP++ Q+AKIIGR  
Sbjct: 130 ASSLGLSYQTMRGIIIPQSTRIALPPLANVLLDLIKASSLAAMITVPPELLQPAK 189

Query: 181 ENDYMSMYILVAFIYWLIALFLERYQFLENKLA 214  
E+DYM+MYIL A IYW I + +Q LE K A  
Sbjct: 190 EFDYMTMYILTAIYWAICSIARVFNILEKKYA 223

An alignment of the GAS and GBS proteins is shown below.

Identities = 110/213 (51%), Positives = 156/213 (72%)

Query: 7 IFNLEKAVKAPPSVIQGLPYTIGLSLVGFILGAIVGFFVAMKMSHPRLLRYLANIHISL 66  
+ N+ L + V+ GLPYT+G+SL+ F G +G +AL+ S L+ YL +IS+  
Sbjct: 1 MINIPMLKDSLGFVLSGLPYTIGISLSFPTGLFGLGLALLGRSROPLIHLYVRAYISI 60

-1726-

Query: 67 MRGIPALVFLFLYGLPLFGIQLDAVTSIVGFTMSSAYISRIIRAALIAVDHQWEA 126  
 MRG+P++V+LF++YPGLP+ G++L A+ + +GF+M+SAAYISE+ R+++ A+D GQWEA  
 Sbjct: 61 MRGVPMITVVLFLYGLPLPYGLLEPALLCAYLGFSMVSAATYISEVFRSSIEALDKGQWEA 120

Query: 127 ARALGLKTPTIYRGILIPQNTIALPSLNLVLLDMVKSSSLTAMITVPDIFNNAKIVGT 186  
 A+ALGL + + II+PQA RIA+P L NV++DMVKSSSL AMITVPDIF NAKI+GG  
 Sbjct: 121 AKALGLPYALMVKKIILPQAFRIAPVPLGNVITDMVKSSSLAAMITVPDIFQNAKIIGR 180

Query: 187 YSDYMTATIALVALIYVVICLTATICQDMWEKRL 219  
 DFM+ YILVA IYW+I L Q++ E +L  
 Sbjct: 181 EWDYMSMILVAFIYMLIAFLERYQEFLENKL 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1553

A DNA sequence (GBSx1645) was identified in *S. agalactiae* <SEQ ID 4795> which encodes the amino acid sequence <SEQ ID 4796>. Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CB12132 GB:Z99105 similar to amino acid ABC transporter  
 (binding protein) [Bacillus subtilis]  
 Identities = 127/276 (46%), Positives = 183/276 (66%), Gaps = 12/276 (4%)

Query: 3 KTIILLGLVLSAMTLAACS--HQSSKEITWDNIKKDGLVKVATPATLYPTSYDHHK-- 58  
 K++ + LAACS N SK-T W+ IK G+ VAT TLXPTSY+D  
 Sbjct: 8 KAVIFSPTMAFFLILAAACSGHNEADSKDTGMEQIKDKGIIVATSGTLYPTSYHDTDSGS 67

Query: 59 -KLITGYEIDMKAIKKLKKIKVFPVEGVVAESPTSDSGKVDVAVNFDITPERLKKCNF 117  
 KLITGYE++++ AK+L +KV+P E+G+ T+V+SG+VD A N+ D T +R +K+ F  
 Sbjct: 68 DKITGYEVEVREAAKRLGLKVEPKMSGIDGMLTAVNSGQVDAANDIDVTQDREKFAF 127

Query: 118 SQPYKYSVGMIVRADGSSKITAKDLSDMKGKHAAGGAGTQYMKIAKQGGAEPIVDNVT 177  
 S PYKYS G IVR D S I K L D KKKKA G A T YM++A++ GA+ VIYDN T  
 Sbjct: 128 STPYKYSYGTAVIRKDDLSGI--KTLKDLKGGKAAGNATVYMEVARKYGAKEVIYDNAT 185

Query: 178 NDVILRDVSTGRITDFIPNDYVTVQVIAVKYVTKYQYDQIKVM- GDVKYNPTQGIWMSKD 236  
 N+ YL+DV+ GRID I NDY Q +A+ +PD+ + + D+KY P +Q +WM K +  
 Sbjct: 186 NEQYLDKDVANGRTDVIINDYIYQLTAL----AAFDLNTIHDPIKIMFNKALVWMSKN 241

Query: 237 KSLKFKIDAAIKDKMGKDGSLKKISPKYVAGQDLTKE 272  
 +L+ K++ A+K+M KDGSL K+S++++ D++K+  
 Sbjct: 242 AALQKKNNEALKKMSKDGSLTKLSKQFFNKADVSKK 277

There is also homology to SEQ ID 1190.

SEQ ID 4796 (GBS183) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 2; MW 33kDa).

GBS183-His was purified as shown in Figure 199, lane 7.



-1727-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1554

A DNA sequence (GBSx1646) was identified in *S. agalactiae* <SEQ ID 4797> which encodes the amino acid sequence <SEQ ID 4798>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1514 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF09821 GB:AE001885 6-aminohexanoate-cyclic-dimer hydrolase  
 [Deinococcus radiodurans]  
 Identities = 178/488 (36%), Positives = 265/488 (53%), Gaps = 17/488 (3%)

Query: 5 DATAMVQAIKQHKISSQELVEQALYKIEEINUSVNAVVSQKYNARQAAYANESHA--- 61  
 DA + Q ++ ++8+++ AI++ + NV++NAVY Y++ A+ + + A  
 Sbjct: 54 DALDLAQLFRRGELSNEDCTAIHFAQVNVVALNAVYIPLTDQLAQARATDAARARGE 113

Query: 62 ---PFAGVPILLKDLQCNQGLSTSGSLFKHHAKQTDVLYQSFEKGLFILLGRINT 117  
 PFAGVP L+KD G G T G++ ++ + D LV+ ++ G + LG+TNT  
 Sbjct: 114 QATGPFAGVPLVQDFGSLRAGVHTGTRAYRDOIPEWDELVRWQAAGLLFLGKINT 173

Query: 118 PEPQFNINISDQLHONVWLPFDHNRWAGSGGAAAVSGBVPIAGASDGGGSIIRIPAS 177  
 PEP +++ +LHG P+D R GSSGG+A+V+G+V+P+AGA DGGGSIIRIPAS  
 Sbjct: 174 PEFALMGVTEPELHGPTNRFDWLGRTGSSGGASAVANGIVPLAGAGDGGGSIIRIPAS 233

Query: 178 FNLGLGLKPSRGRIPVGFSSYRGWQAASHFALTKSVRDTKRLLYLQSYQVES---PF 233  
 GL GLKPSRGR+P G WQA+ LT+SVRD+ LL Q + P  
 Sbjct: 234 CGGLGLKPSRGRVPOGDGVGEFWQA+VEHVLT+SVRDSALLDLEGGPDAGALFLP+ 293

Query: 234 PLKGLKESLPEFSVSKPLKIAVLMDSPKATKVSSEKAAIKKADPLSKGNHLELVEQ 293  
 P + S+E E L+I PL V E AA++ AA L G+ +E V  
 Sbjct: 294 PERPYSEEVGRE---PQLRIGFSTANPLGRVHPECVA+VQAARALLES+LGEVVEVAL 350

Query: 294 PLDGIHSMKTYCHMSVETAMPDFDIKSLGRSMFSDMELATWAMYQSGQRVLAKDYK 353  
 P DG ++ + M+ ET A + +LGR SD+E +TW + Q+ G+ A D++  
 Sbjct: 351 PWDGPAQAQPLMLYPGETGASLAALRDTLGRPARASDV+AVTWLQQLGRSYSAADFAA 410

Query: 354 LLDSDQPAATMARFHEYNLDILTAINTQFADPHQGD---LDETQKOLRHMGEFSVSE 410  
 SW+ A M RPH+NYDL+LT P G+ + L + M + +  
 Sbjct: 411 ARASVNVHARAMGRPHQNYDILLTFVPLATPFIQIGELQPRVQAALRLAAQMDVSGLLR 470

Query: 411 QQDLIKWMFEDSMWTPPTTHQPNLTQPSLAIPHTLTKKGLPLGVQLTAAKGRREDLLAV 470  
 + + + D + P+T NLTPQ++P+P H T +GLP+GWQ A RED+LL+  
 Sbjct: 471 RSGVDALATDILEKMPYQLANLTQPFAMSVPLHWADGLPVGVQFVAPLAREDVLLRL 530

Query: 471 AELFEKEK 478  
 A E+ +  
 Sbjct: 531 AGQLQAR 538

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4047> which encodes the amino acid sequence <SEQ ID 4048>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

-1728-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 277/484 (57%), Positives = 348/484 (71%), Gaps = 2/484 (0%)

Query: 1 MVFKDATAMVQAIKQHKISSQELVBOAIYKIRQNVSVNAVSKQYNEARQAQYANESN 60  
 M ++DATAM A++ + + ELV QAIYK ++ N ++NA+ S+++ A + AK + S  
 10 Sbjct: 1 MTYQDATAMAIJAVQTQQTPLLELVTOAIYKAIKLNPTIANITSERPEANLEAKQORDPSG 60

Query: 61 APFAGVPILMKDLQGNQKQJSTSGSQLPKHYHAKQTDYLVQSFELGFIILGRNTIPEF 120  
 PFAGVP+ LKDLQK KG STSGS+LPK Y A +TD V+ E LGFIILGR+NTIPEF  
 15 Sbjct: 61 LPFAGVPLFLKDLQELKGHSSTSGSRLPFKEYQRTKTDLVKRLGALGFIIILGRSNTIPEF 120

Query: 121 GPKNISDGQLHGNVNLPPDHSRNAGSSGGAAAAVSSGMVPIAGASDGGGIRIPASFNG 180  
 GPKNISD LHG VNL P D++RNAGSSGGAAA VSSG+ +A ASDGGGIRIPASFNG  
 Sbjct: 121 GPKNISDSLSHGPNLPDRTNRNAGSSGGAAALVSSGTSALATASDGGGIRIPASFNG 180

Query: 241 ESLEPFSVSKPLKIAVLMDSPKTKVSSEAKAIKEADPLSQKNHL-ELVQPLDGIH 299  
 +S+++ S+ +PL IA + VS + A+++A +L +G+ L EL E P++  
 25 Sbjct: 241 DSITQ-SLQRPLTIAFYQRLSDGSPVSLDTAKALRQAVTLWRBQHLVLEBPFFNMTE 299

Query: 300 SHKTYCMNSVETAMPDIEKSLGRSMEFSDMELMTWAMYQSGQVLAQDYSKLSD 359  
 ++ Y +MNSVETAMP DIE + GR M DME MTWA+YQSG+ A YS+L L WD  
 30 Sbjct: 300 VIRHYIMNSVETAMPADIETGRPMTKDDMETMTWAIYQSGQDPAWKYSQVQLKWD 359

Query: 360 QFAATMARPHENYDLILTAATNQAPFPHGQFDLDETQKQLRHWSGSPVSEQDLIWKMP 419  
 ++ATMA FHE YDL+LT TN PAP HG+ D L L FS EQ +L+ MP  
 35 Sbjct: 360 TYSATMASPHETIDLLTFTTNTTAPKHGELVPSDKIMANLAQAPISSESEQFNLVETMP 419

Query: 420 EDSMAWTFPTHQPNLTGQPSLAIPHTLTKEGLPLGVLTAAKGREDLLAVARLFEKEQ 479  
 S+A P+T PNLTGQP++++PT+ TKEGL +G+QL AAGREDLLL +AE FE  
 Sbjct: 420 GKSLAINPYTALPNLTGQPAISLFTYETKEGLSMGIQLIAAKGREDLLIGTAEQFEANGL 479

Query: 480 FKGP 483  
 K P  
 40 Sbjct: 480 LKIP 483

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 45 vaccines or diagnostics.

### Example 1555

A DNA sequence (GBSx1647) was identified in *S.agalactiae* <SEQ ID 4799> which encodes the amino acid sequence <SEQ ID 4800>. This protein is predicted to be transcription elongation factor (greA). Analysis of this protein sequence reveals the following:

50 Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5003 (Affirmative) < succ>  
 55 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14674 GB:Z99117 transcription elongation factor [Bacillus subtilis]  
 60 Identities = 86/154 (55%), Positives = 114/154 (73%), Gaps = 1/154 (0%)

-1729-

Query: 3 EKTYPMVQVEKDQLEKELEELKLVRRPEVVERIKIARSYGLSENSEYDAAKBQAFV 62  
 EK +PMV K +LE+ELE LK V+R EVVERIKIARS+GLSENSEYD+AK+BQAFV 62  
 Sbjet: 4 EKVPMTABGQKLEKELEELKLVRRPEVVERIKIARSYGLSENSEYDAKESQAFV 63

Query: 63 QIQILETKIRYAEIIDSDAVADEVAIGKTVLVQEWGNTDKDTYHIVGAAGADIFSGKIS 122  
 ++ LE IR A+II+ D + V +GKTV E+ D+++Y IVG+A AD F GKIS 122  
 Sbjet: 64 RVTITLNMIRNAKIIEDDG -GENVGLGKTVFVLPDGDSESYTIVGSARADPFSGKIS 122

Query: 123 NESPIAHALIGKKTGDLATIESPAGSYQVEIISV 156  
 N+SPIA +L+GKK + T+++G V+I+ +  
 Sbjet: 123 NDSPIAKSLGKKVDEEVTVQTPGGHMLVKIVKI 156

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4801> which encodes the amino acid sequence <SEQ ID 4802>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4434 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/160 (90%), Positives = 149/160 (92%)

Query: 1 MAEKTYPMVQVEKDQLEKELEELKLVRRPEVVERIKIARSYGLSENSEYDAAKDEQAFV 60  
 MAEKTYPMV EK+CLEKELEELKLVRRPE+VERIKIARSYGLSENSEYDAAKDEQAFV 60  
 Sbjet: 17 MAEKTYPMVLEKEKELEELKLVRRPEVVERIKIARSYGLSENSEYDAAKDEQAFV 76

Query: 61 EQQIQILETKIRYAEIIDSDAVADEVAIGKTVLVQEWGNTDKDTYHIVGAAGADIFSGK 120  
 EQQI LETKIRYAEIIDSDAVADEVAIGKTV+VQEWGT DKDTYHIVGAAGADIFSGK 120  
 Sbjet: 77 EQQISTLETKIRYAEIIDSDAVADEVAIGKTVLVQEWGNTDKDTYHIVGAAGADIFSGK 136

Query: 121 ISNESPIAHALIGKKTGDLATIESPAGSYQVEIISVEKTN 160  
 ISNESPIA ALIGKKTGD IESPA +Y VEIISVEKTN 160  
 Sbjet: 137 ISNESPIAALIGKKTGDKVRIEPAATYVEIISVEKTN 176

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1556

A DNA sequence (GBSx1648) was identified in *S.agalactiae* <SEQ ID 4803> which encodes the amino acid sequence <SEQ ID 4804>. This protein is predicted to be aminodeoxychorismate lyase-like protein. Analysis of this protein sequence reveals the following:

possible site: 58  
 >>> Seems to have no N-terminal signal sequence

INTRORAL Likelihood =-13.64 Transmembrane 238 - 254 ( 230 - 260)

----- Final Results -----  
 bacterial membrane --- Certainty=0.6456 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein  
 [Streptococcus thermophilus]  
 Identities = 135/210 (64%), Positives = 171/210 (81%)

-1730-

Query: 373 KTTSTPYKADDFLKLVDQSTFIKRMVAKYPNLLGSLDKSKAIYQLSGYLFPATYMYKD 432  
 K +ST K DFLKI++D+ FI KM AKYP Lb +Lp+ + A Y LSGYLFATYIN + D  
 Sbjct: 5 KHSSTGLKEDFLKLAKDDAFITMKAKYPTLLANLPSNDAKYVLESGYLFPATYINLDD 64

5 Query: 433 TTLBGLVEMDMISMTNKMAPPYNTIKAKNMSVNDVLTLSLIVEKEGSDTDDRRIASVFPY 492  
 TT+E L E+M+ TM+T ++PYY TI + N +VM++LTL+SLVEKEG+TD+DR+ IASVFPY  
 Sbjct: 65 TTVESLAREMLPMTDHLSPYYATILSGMHNVNELLTLASLVEKEGATDDDRKNIASVFPY 124

10 Query: 493 NRLGAGQALQSNIAILYAMGKLGDKTSLAERDAQINTSIKSPYNIYNTGLMPGVDSPSI 552  
 NRL++ ALQSNI+LY +GKLG +T+L EDA I+T+I SPIN Y + GLMPGVDSPS+  
 Sbjct: 125 NRIASDMALQSNIADVILVGLKLGQETTLKEDATIDTINIDSPINDYVHKLMPGVDSPSL 184

Query: 553 SAIEATIKPASTDYLYFVADVKIKGVYAK 582  
 SAIEA I P+ST Y+YFVADV IGRVY+A+  
 Sbjct: 185 SAIEAVINPSTKMYFVADVSTGRVYFAE 214

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4805> which encodes the amino acid sequence <SEQ ID 4806>. Analysis of this protein sequence reveals the following:

Possible site: 59

20 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.91 Transmembrane 161 - 177 ( 155 - 183)

----- Final Results -----

25 bacterial membrane --- Certainty=0.4163 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

30 >GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein  
 [Streptococcus thermophilus]  
 Identities = 135/212 (63%), Positives = 161/212 (75%)

35 Query: 295 KTKKAKTPFNEKDFLDVLDTDAFIQDMVKRYPKLLATIPTKKAITYRLSGYLFPATYNY 354  
 K K + T EKDF L+ D+AFI N +YP LLA +P A Y LSGYLFATYIN +  
 Sbjct: 3 KGRHSSTGLKEDFLKLHDDAFITMKAKYPTLLANLPSNDAKYVLESGYLFPATYINLH 62

Query: 355 KETTMRELVEDMLAANDATLVYYDKIAASGKTVNEVLTLASLVEKEGSDTDDRRIASV 414  
 +TT+ L E+ML MD L PYY I +S VNE+LTLASLVEKEG+TDDR+ IASV  
 40 Sbjct: 63 DTTVESLAREMLPMTDHLSPYYATILSGMHNVNELLTLASLVEKEGATDDDRKNIASV 122

Query: 415 FYNRLNSMALQSNIAILYAMGKLGDKTSLAERDAQINTSIKSPYNIYNTGLMPGVASS 474  
 FYNRLNS MALQSNI+LY +GKLG +TTL EDATIDT I+SPYN Y + GLMPGV S  
 45 Sbjct: 123 FYNRLNSMALQSNIADVILVGLKLGQETTLKEDATIDTINIDSPYNDYVHKLMPGVDSF 182

Query: 475 GSAIEATINPASTDYLYFVADVITGVYAK 506  
 +SAIEA +NP+ST Y+YFVA+V TG VY+A+  
 Sbjct: 183 SLGAIEAVINPSTKMYFVADVSTGRVYFAE 214

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 310/603 (51%), Positives = 403/603 (66%), Gaps = 86/603 (14%)

Query: 1 MTFPNDDQHSNHDQKSFKEQILAELEERANRLRKLEELLYQKEQBAKAAARRTAQLMADY 60  
 +T+FK D + Q+SFKEQILAELE+AN+RK +EELI+  
 55 Sbjct: 3 LTFDFKDQQDQ-QRSFKEQILAELEKANQIRKEKEELF----- 41

Query: 61 EAQRLKDEREAREAKALETQRIEBQKARIBAKLLAEAREERERRQAEQALASQEQVIN 120  
 ++ LE +E AR A+L AE R++ A Q+E + +  
 60 Sbjct: 42 -----QKRLAKKAARRTAQLYARYKRD-----APKQESIAH 74

Query: 121 QGMPSRELDGSGSKSEFPRTTENVPDIDLKADKIDVATAVNPQETEEI FLVRATDIPTEG 180  
 +T ++ +A K V T+ + T +E  
 Sbjct: 75 NN-----KPAK-----PQAKGAVMTSEALKPT-----LLESK 103

-1731-

Query: 181 ENVKLGISELSFVAKPIRVEDLSKEEGIALSAKNNKRR---RQADNVAKRIAR 237  
 EN L ++ A E +++ +E +L+ +H+ R + RQ+ + AK+L+  
 Sbjct: 104 ENSSLATTKKRVQANE---LQETASKEQVPLTIEKGHSVRRLSKRQOTERAACKIST 160

5 Query: 238 ILISIIILVLLJTFAPVGYRFVDSAIKPVDSNKNFVQVEIPQSGNKLIGQILEKGVK 297  
 +LIS II+ LL G +V SA+ PVD NS+ FVQVEIP GSGNKLIGQII+K G+IK  
 Sbjct: 161 VLISIIITLLAVLAGGYVYSALNPVDKNSDAFVQVEIPSGSGNKLIGQILKKGLIK 220

10 Query: 298 SATVFNYSKFNQYSNFGSGYYNLKKGMTLDQIAAELEKGGTAEPKPAKGKILITRYT 357  
 ++TVF++Y+KFN++NFGSGYYNL+KSN+L++IA+ L++GGTAEPKPA+KGLILI EGYT  
 Sbjct: 221 NSTVFSFYTKFNFTNFGSGYYNLKQSNLSLEIASALQGGTAEPKPAKGLILIPGYT 280

15 Query: 358 IKQIAKAIASN-KIDTKTISTPYKADDFLKLVDQETFIKKNVAKYFNLGLSLEPKSKAIY 416  
 IKQIAKA-E N K TK TP+ DEL LV DE FI+ MV +TP LL ++P K KAIY  
 Sbjct: 281 IKQIAKAVEHNSKGTKKAKTPFNKDFLDLVTDGAFIQDMVKRYPKLATIPTKKEAIY 340

20 Query: 417 QLEGYLPATINYYKOTTLBGLVDMISTMTNWKAPYYNTIKAKMSVNDVLTLSSLEK 476  
 +LEGYLPATINYYK+IT+ LVEDM++ M+ + PTY+ I A +VN+VLTL+SLVEK  
 Sbjct: 341 RLEGYLPATINYYKETIMRELVEDMLAAMDATLVPIYDKIAASGKTVMNEVLTLASLEK 400

25 Query: 477 EGSTDDEDRKIASVFNNLSAQALQSNHAILYAMKLGDKTSLAEDAQINTSIKSPYNI 536  
 EGSTD+DRK+IASVFNNRL++G ALQSNHAILYAMKLG+KT+LAEDA I+T+I SPYNI  
 Sbjct: 401 EGSTDDEDRKIASVFNNRLNSGVALQSNHAILYAMKLGDKTSLAEDATIDTTINSYNI 460

30 Query: 537 YNTGLMPGPFVSSGVSIAEATKNPASTDYLYFVADVKTNVYAKDFETHKANVEKYIN 596  
 YNTGLMPGPV S +SAIEAT+ PASTDYLYFVA+V TO VYAK FE H ANVEK+N  
 Sbjct: 461 YNTGLMPGPFVASSGVSIAEATLNPASTDYLYFVAVHVGVEVYAKTFEHSANVEKYIN 520

Query: 597 SQI 599  
 SQI  
 Sbjct: 521 SQI 523

A related GBS gene <SEQ ID 8843> and protein <SEQ ID 8844> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 8  
 McG: Discrim Score: -17.88  
 GVH: Signal Score (-7.5): -3.51  
 Possible site: 58  
 >>> Seems to have no N-terminal signal sequence

40 ALOM program count: 1 value: -13.64 threshold: 0.0  
 INTEGRAL Likelihood = -13.64 Transmembrane 238 - 254 ( 230 - 260)  
 PERIPHERAL Likelihood = 5.78 285  
 modified ALOM score: 3.23

45 \*\*\* Reasching Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.6456 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00931(1417 - 2046 of 2400)  
 GP[8574530]gb|AA77615.1|AF151720.1|AF151720(5 - 214 of 214) aminodeoxychorismate lyase-  
 55 like protein (Streptococcus thermophilus)  
 %Match = 17.5  
 %Identity = 64.3 %Similarity = 81.4  
 Matches = 135 Mismatches = 39 Conservative Sub.s = 36

60 1236 1266 1296 1326 1356 1386 1416 1446  
 NYTSKFNYSNFGSGYYNLKKGMTLDQIAAELEKGGTAEPKPAKGKILITRYTQIAKAIASNKIDTKTISTPYKAD  
 | : | : |  
 AKKGKHSSTGLKEK  
 10

-1732-

[illegible]

SEQ ID 8844 (GBS370) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 6; MW 70kDa).

20 GBS370-His was purified as shown in Figure 209, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1557

A DNA sequence (GBSx1649) was identified in *S.agalactiae* <SEQ ID 4807> which encodes the amino acid sequence <SEQ ID 4808>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0183(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10077> which encodes amino acid sequence <SEQ ID 10078> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA98689 GB:Z74367 ORF YDR071c [*Saccharomyces cerevisiae*]  
Identities = 52/174 (29%), Positives = 81/174 (45%), Gaps = 18/174 (10%)

40	Query: 27	MSMIIRLQDLQVLSIQISFSSAAEAGAKQMRUITMITDIT--FLVAETNR-- 80 + M I R +E D L +Q + + + + S F E A S + + R L + + E I G A	
	Subject: 10	LHWYIRPLILRLDLKQLINLESQGGPPNRRSEIISFLRLNCPKISGLFIRITEGKEV 69	
45	Query: 81	--LASYIVKPKYKRYITLDFKIFVSEFFPRVGQFISITSLSIHPDFQGGISATLLA 137 L G I G I Y T + K V + W I G I S I P + + + + T L L L	
	Subject: 70	KETLIGHMIGKTLPHRYTITSMKQIQ--VESSNHSIHSSVVFYKQKNLATTLLD 126	
	Query: 138	MDLVVSVN--RDSGISLTCHDLSISFYFMGPKRDEGS--DSKHGSLWYNM 185 + Q E + I L H + I L F Y E G P K E + D K W + M	
50	Subject: 127	YIQKSNISQKNIKVLHLAEPLIFPKYKQKGLIARNTNADKDNKFAQWIM 180	

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4809> which encodes the amino acid sequence <SEQ ID 4810>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

-1733-

## ----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.2576 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/159 (54%), Positives = 117/159 (72%), Gaps = 1/159 (0%)

10   Query: 29 MIIRNCCLEDLQCVISIEQINFSEARAAKAKMQRITIMIDTFLVAERNGRLAGYIEGP 88  
           M+IR   DL+ + +IE NFS EA ++ ++S + ++ DTFLLVA I+ + GYIEGP  
       Sbjct: 1 MLIRQVQSDLEVIATIESDNFSPQEAITRAVLEEHRILIPDTFLVALIDQETVGYIEGP 60

15   Query: 89 VIKGRYLTDLDFHIVSEFFRVGGPIGITSLSIHDFKQGGIGTALLAAKDLVVSQERD 148  
           V+   L D LPH V++ F + GG+I ITSLSI F+ QG+GTALLAA+KDLVV+Q+R  
       Sbjct: 61 VVTFEILEDLPHGVTKNF-KTGGYIAITSLSIKAKHQCCGVGTALLAAKDLVAVQQR 119

20   Query: 149 GISLTHCHDLISFYEMNGFKDEGESDSKGGSLWYNNMIW 187  
           G+ LTCHD LIS+YEMNGF ++G S+S+GG+LWY NNMIW  
       Sbjct: 120 GLILTHCHDLISFYEMNGFINGISBSQGGTLWYNNMIW 158

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1558

25   A DNA sequence (GBSx1650) was identified in *S. agalactiae* <SEQ ID 4811> which encodes the amino acid sequence <SEQ ID 4812>. This protein is predicted to be udp-n-acetylmuramate--alanine ligase (murC/ddIA). Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

30   INTEGRAL   Likelihood = -2.60   Transmembrane   272 - 288 ( 270 - 288 )

## ----- Final Results -----

35           bacterial membrane --- Certainty=0.2041 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC00294 GB: AF008220 putative UDP-N-acetylmuramate-alanine  
           ligase [Bacillus subtilis]  
       Identities = 238/432 (55%), Positives = 315/432 (72%), Gaps = 3/432 (0%)

40   Query: 5 YHFIGIKGGMSALALALHQMGEHIVQGSDDVDKYYFTORGLEQAGVTLFPENNLSEDLS 64  
           YHF+GIKG+GMS LA +LH G+ VQGSDD+K+ PTO LE+ +TILFPE NI +  
       Sbjct: 4 YHFPVIGIKGTGMSPLAQTLHDNGYIVQGSDDIEKFIPTOTALREKNNITLFPSENNIKPGMT 63

45   Query: 65 ILAGNAFDPDNNELAYVIEKGYQFKRYHFLGDFMFOPTSLGVAGAHGKISTTGLLAHV 124  
           +IAGNAF FD +E+ + +G RTH+FLGDH++PTS+ V GAHKISTTGLLAHV  
       Sbjct: 64 VIAGNAF-PDTHFEIEKAMSEGISPIRVIRHFKPLGDYMKKPTSAVIGAHGKISTTGLLAHV 122

50   Query: 125 LKNITDTSFLIGDGTGRGSANANVFPVFADEYERHMFYHPEYSIITNIDFHDYFTGL 184  
           ++N TSFLIGDGTG+G+ N+ YFVFEA KY RHP+ Y P+Y+I+TNIDFHDYFTGL  
       Sbjct: 123 IQNAKPTSFLIGDGTGQGNSEYFPVFADEYERHMFYHPEYSIITNIDFHDYFTGL 182

55   Query: 185 EDVFNAPNDYAKCVOKLFIYGEDPKLHEITSEAPIYYGPDSENDFIKIDITRTVNGSD 244  
           +DVF+AF +A QV KG+ G+D L +I + P+ YYG +NDP A+I+ ++ G+  
       Sbjct: 183 DDVFDAPQEMALQVNGKSIACGDEHLKPIHANVFPVYVYTGSENDPQARNIVKSTEGTT 242

60   Query: 245 FKVFYVQKEIGQPHVPAYGKINILNATAVIANLYIMGIDMALVAEHLKTFSGVKRRFTEK 304  
           F VF   P++PAYG HN+IN+ AVTA + ID +++ LK+G GVKRRF EK  
       Sbjct: 243 PDVFRNTFYDTFFYIPAYGHEVNLNSLAVIALCHYEIDSSIIKHALKSGFGVKRRFNEK 302

-1734-

Query: 305 IIDDTVIIDDFAHHPTRIIATLDAARQKYPKSKRIVAIQPHITFTRTIALDLDFEFAHLSQA 364  
 + D V+IDD+AHHPTRI T++AARQKYP +EIVA+PQPHITFTRT LDEFA +LS A  
 Sbjct: 303 QIGDQVLIDDVAHHPTRIKVTITDAARQKYPDRRIIVAVQPHITFTRTQQLDFEFAEISLSGA 362

Query: 365 DSVYLAQIYGSAREVDKGEVVKVEDLAAKIVKHSIDVTVENVSPELLHNNVAVYFMGAGDI 424  
 D VYL I+GSARE + G++ + DL KI ++ Lt ++ S L HD AV +FMGAGDI  
 Sbjct: 363 DCVYLCDFGSARE-NAGKI/TIGDLQKKI -HNKLIERDDTSVLKAKHKAHVILFMGAGDI 420

Query: 425 QLYERSFEELLA 436  
 Q Y R++S ++A  
 Sbjct: 421 QKYMAYENVMA 432

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4813> which encodes the amino acid sequence <SEQ ID 4814>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -4.57 Transmembrane 271 - 287 ( 269 - 288)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2826 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA00294 GB:AF008220 putative UDP-N-acetylmutamate-alanine  
 ligase [Bacillus subtilis]  
 Identities = 236/431 (54%), Positives = 310/431 (71%), Gaps = 2/431 (0%)

Query: 5 YHFIGIKSGMSALALMLHGKHVQGSDEKYYPTQGLQAGITITLPPSEDNITPDME 64  
 YHF+GIKG+GMS LA +LH G+ VQGSDEK+ PTQ LE+ ITLPPS +NI P M  
 Sbjct: 4 YHFIGIKSGMSFLAQLIHNDGYTVQGSDEKFIPTQTALEKKNITLPPSABNIPKMT 63

Query: 65 LIVGNAPRENKKEVAYALRHQIPFKRYHDFLDGFMKSFISFAVAGAKTKSTTGLLSHLV 124  
 +I GNAP + + E+ A+ IP RYH PLGD+MK F S AV GAKTKSTTGLL+EV+  
 Sbjct: 64 VIAGNAPFDTHPEIKAMSEGIPIRYHKLFDGTMKFTSVAVTAGTKSTTGLLAHV 123

Query: 125 KNITDTSYLIQDGTGSGSANAQYFVFSEDEYERHFMPIYHPSYIITNIDFDMFDYPTGIA 184  
 +N TS+LIGDGTG+G+ N++YFVFE+ EY RHF+ Y P+Y+I+TNIDFDMFDYF+ I  
 Sbjct: 124 QNAKPTSLIQDGTGCGENSEYFVFCEYRRHFLSYQPDYAINITNIDFDMFDYFSSID 183

Query: 185 DVNAPNDYAKVKKALFVYGEDDELKKIEAPATIIYFGFSEGNDFIADITKTGTSDF 244  
 EV +AF + A QV K+ G+D+ L KI A P+ YYG E NDF A +I ++T G+ F  
 Sbjct: 184 DVPDFAFGMALQVNGKIIACGDDEHLFKIHANVPVYVYGTGEENDFQARNIVKSTGTF 243

Query: 245 KVYHQGEVIGQPHVAYGKHNLINATV+ANLTVAGIDMALVADHLKTSVGRKRRFTEKI 304  
 V + F++PAYG HN+LN+ AVIA ID +++ LK+F GVKRRF EK  
 Sbjct: 244 DVPVRNTFYDFTIIPAYGHHNVNLSLAVIALCHYREIDSSIIKHALSPGKVRKRRFNRKQ 303

Query: 305 INDVTIIDDFAHHPTRIIVATDAARQKYPKSKRIVAIQPHITFTRTIALDLDFACALNSAD 364  
 + D ++IDD+AHHPTRI TI+AARQKYP +EIVA+PQPHITFTRT L++FA +L+ AD  
 Sbjct: 304 IGDQVLIDDVAHHPTRIKVTITDAARQKYPDRRIIVAVQPHITFTRTQQLDFEFAEISLSGA 363

Query: 365 SVYLAQIYGSAREVDKGEVVKVEDLAAKIKPSQVVTVENVSPELLHNNVAVYFMGAGDIQ 424  
 VYL I+GSARE + G++ + DL K I +++ ++ S L HD AV +FMGAGDIQ  
 Sbjct: 364 DCVYLCDFGSARE-NAGKI/TIGDLQKKI -HNKLIERDDTSVLKAKHKAHVILFMGAGDIQ 421

Query: 425 LYERSFEELLA 435  
 Y ++E ++A  
 Sbjct: 422 KYMAYENVMA 432

An alignment of the GAS and GBS proteins is shown below.

Identities = 369/443 (83%), Positives = 406/443 (91%), Gaps = 1/443 (0%)



5	Query: 61	L E I I A G N A P P D N E S L A B I V K Y G Q F K R Y E H F L D G M P R Q T S I G V A G A R G K T S T T G L	120
		D-E-I GNAFR +NN+R+AY + FKRYH+FLDGM+ S S VAGAGRGKTSSTTGL	
	Sbjct: 61	P D E I L V G N A P R F N K R H A Y A L A H Q I P K R Y H D F L D G M P R Q T S I G V A G A R G K T S T T G L	119
10	Query: 121	L A H V L K N I T D T S F L I G D T G R G S A N A N Y V F P E A D K Y E R H M P M Y H K Y S I I T N I D F M F O Y	180
		L+HVLKNI+TDS+LIGD+TGRGSANAY+VPFPEADKYERHMP+MYHYSIITNIDFMFOY	
	Sbjct: 120	L H V L K N I T D T S I L G D T G R G S A N A Q Y V F P E S D E Y E R H M P M Y H P E S I I T N I D F M F O Y	179
15	Query: 181	F T G L E D V F A P N E Y A Q Q D G K S F T Y G E D P K L H E I T S E A P I Y Y Y G F E D S N D F I A K D I T R T V	240
		FTG- DV NARNYAYQ+Q+L+FYGED+L+I+ APIYYGGE+ NDFIA DITRT	
	Sbjct: 180	F T G L A D V A P N E Y A Q K V A L F Y G S D E L K I K A F A P I Y Y Y G F E D N F I A D I T R T V	239
20	Query: 241	N G S D F K V Y N G E I G Q F H V P A Y G K N I L N A T A V I A N I Y N G I D M A L V A H E I L T F S V G K R R	300
		NGSDFK+ E IGFQFHVPAYGKNTLNATAVIAN+ GIDMALVA+HKTFSVGR	
	Sbjct: 240	N G S D F K V H G E V I G Q F H V P A Y G K N I L N A T A V I A N F V A G I D M A L V A H E I L T F S V G K R R	299
25	Query: 301	P T E K I I D D I V I D D F A H P P E T I E A L T D A A R Q K Y P S K E I V A I P Q H F T R T I A L L D E F A H A	360
		PTEKII+DT+IIDDFAHPPET+IAT+DAARQKYPSEIVAI+PQHFTRTIALD++FAA	
	Sbjct: 300	P T E K I I D D I V I D D F A H P P E T I E A T D A A R Q K Y P S K E I V A I P Q H F T R T I A L L D E F A H A	359
30	Query: 421	L S Q A D S V Y L A Q I Y G S A R E V D N G E V K V E D L A A K I V K H S D L V T V E N V S P L L A N D H N A V Y F M G	420
		L++ADSVLYAQIYGSAREVDN+GEVKVEDLAAKI+K+S+VTYVENSPLL+HNNAVYFMG	
	Sbjct: 420	L N E A D S V Y L A Q I Y G S A R E V D N G E V K V E D L A A K I K P S Q V T V E N V S P L L D H N A V Y F M G	419
35	Query: 421	A G D I Q L Y E R S F E E L L A N L T N I Q 443	
		AGDIQLYSEFEELLANTNQ	
	Sbjct: 420	A G D I Q L Y E S F E E L L A N L T N Q 442	

35 product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 8; MW 74kDa), Figure 33 (lane 8; MW 74kDa) and Figure 37 (lane 3; MW 74kDa).

40 protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 4812 (GBS157) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 11-13; MW 74kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45     **Example 1559**

A DNA sequence (GBSx1651) was identified in *S.agalactiae* <SEQ ID 4815> which encodes the amino acid sequence <SEQ ID 4816>. Analysis of this protein sequence reveals the following:

```

50      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1980(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1736-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4817> which encodes the amino acid sequence <SEQ ID 4818>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2731(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 80/201 (39%), Positives = 126/201 (61%), Gaps = 9/201 (4%)

Query: 7  RFPLIADDEFVMSPLVQNLVDNEDLNNIRDFYQKTYQSMVKSNEYHEISHPKVIEN 66
      +FPI+AD + P +M LY+NEDLI NTR +YQ+K Y + ++ ME +
Sbjct: 5  QFPLVADGTAISDPAKQALYENEDLITNIRGTYQDKYDDIARN----EEPTAKATSRQ 60

Query: 67  DVPVPPQ--SFVKGATLSEFSEQKAKRSVREKRCAYTAQSEFKAPSKAEAPQQQLKATVPEKK 124
      P + S +K + ++RQ+AK+ ++EKRCAY AK+ P + + +QQ + P +
Sbjct: 61  TFSSKRFCSENDKGHVYVKEARQKAKQDLKEKRCAYLAKEMAYVFKQVSKKQPADSSPSQ 120

Query: 125  QTORVKTLSHLSDRLQOBSYILAEPIIIPCEPDNTPNP-KTKQINPDLAKRSQVYNNKQD 183
      + + TS+S + +L Q+YILAE+P ++EP N P TKQIN+DPLK SQ+YN ++
Sbjct: 121  X--QATTENSRFTKLLHQDNVILARLPKEYKEPKQLFQQGTTKKQNYDLKSSQITNNKE 178

Query: 184  NQPHKERAKAQSLNLTRFKDI 204
      + +E+ AQSLNL+RF+D+
Sbjct: 179  MRQQRKTTAQSLNLTRFEDL 199

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1560

A DNA sequence (GBSx1652) was identified in *S.agalactiae* <SEQ ID 4819> which encodes the amino acid sequence <SEQ ID 4820>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4959(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1561

A DNA sequence (GBSx1653) was identified in *S.agalactiae* <SEQ ID 4821> which encodes the amino acid sequence <SEQ ID 4822>. This protein is predicted to be SNF. Analysis of this protein sequence reveals the following:

-1737-

Possible site: 28

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.32 Transmembrane 743 - 759 ( 743 - 759)

## 5 ----- Final Results -----

bacterial membrane --- Certainty=0.1128 (Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

## 10 The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:CA67095 GB:X98455 SNF [Bacillus cereus]

Identities = 259/678 (38%), Positives = 406/678 (59%), Gaps = 21/678 (3%)

Query: 369 QNEILLQMVFDYGNLTVINRQELBQLTFASHFKHEKVKFKLEKYGFAPHFSTSHAYS 428  
 15 +N +L + F YGN + ++ + F K E+++ ++ + FA + ++  
 Sbjct: 388 KNRLLAGLEPHYGNVVFLEEDQPSVFNREDEKKEILDIMSESAFAKT-BGGYPMHN 446

Query: 429 AQELYDFYTYMLPQFKKQMTV--SLSAKLESTYRIERQIDIEAKGSL--LDISDFDSLD 484  
 20 + Y+F +++P K + + + + KL + + P I + K + L FD +  
 Sbjct: 447 EFAYNYFLYHYVPTLKGVLDIYATTAIKLIRHSGDIAPLIRVRKERIDWLISFRDITKGI 506

Query: 485 LENDVDQALVALFDNNFYFVNKSGQLVIFD-EETKKVSATLQ--GLRARRAKNGHIELDN 541  
 E ++ L AL + Y + +G L + +E +++ ++ G+R + +  
 Sbjct: 507 FEAEIKGVAAALEEKRYKYLKANGSLLSLESEKFEINQFVKESGIRKEFLHSGEVNVPL 566

Query: 542 IAAFQLSELFAQDNVSFSQHFYQLIEDLRHPEKFK--IPGLSVSASLRDYQITGVNRLS 599  
 I + + + +S + +L+E +++P+K K +P ++ A +R+YQ+ G W+  
 Sbjct: 567 IRSVMKXGLHGEHGVLSLDESVCGLVESIQNFKILKFTVPP-TLHAVMREYQVYGFEWK 625

Query: 600 MLDHYGFAGILADDMLGKTLTQITISFLSKLT--RDSR--VLILSPSSLIYNKQDEPHFK 655  
 30 L +Y F GILADDMLGKTLQ+I+++ + L R+ + +L+SPSSL+YNN E KF  
 Sbjct: 626 TLAYYRFGSILADDMLGKTLQSIAYIDSLVPEIREKKLPILVSPSSLVNVMSEKLFK 685

Query: 656 APDVIDVAVYSGKIRDEIIAE--RQVITITSYSSFRQDFSTYSEGNVYDILBEAQVMK 713  
 35 AP + +A G++ R +I+ + V+ITSY R+D +Y+ + L LDEAQ K  
 Sbjct: 686 APHRAVIADGNQTERKKILKDVARFDVITSYLLARDFVRSYARP-FHTLFLDEAQAFK 744

Query: 714 NAQTKIAHSLRSFEVKNCFALSGTPIENKLLINSIFQIILPGLLPGRKEFLKLNPKQVA 773  
 N T+ A +++++ + F L+GTP+EN L E+NSIP ++ P LLPG+KEF L + +A  
 40 Sbjct: 745 NPTQTARAVKTIQASYRFGLTGTGVENSLEELNSI FHVVFPELLGRKEFGDLREDA 804

Query: 774 RYIKPFVMRRRKSEVLPELPLDIENFYFNEMTDSQVIYLAQLRQI-QESIQHSSDADEN 832  
 +KPFV+RR KE+VL ELFD IE +E+ QK +Y R L ++ +E+++H L  
 45 Sbjct: 805 NAVKPFVRLRLKEDVLQELPDKTEHLQSSLELFPDQRLRYAAYLAKLREETLKHLDKTLR 864

Query: 833 RRKIELSGITRLRQICDTPRLFMD-YDGSQLESLSQLLTQIKENGHRALIFSQFRGM 891  
 + KI IL+G+TRLRQIC+ P LF+D Y G S KLE L +L + + G R LIFSQF N  
 Sbjct: 865 KKKIRILAGITRLRQICNHPALFVDYDYGSSAKLEQLDILEECRSTGRILIFSQFTKM 924

Query: 892 LDIAEREMVAMGLTYYKITGSTPANRHEHMITRANAGSKDAVLLSLKAGGVGINLTGADT 951  
 L I RE+ + + +G+TP+ ER E+ FN G D FLISLKAGG GINLTGADT  
 50 Sbjct: 925 LSTIGRELNRQAIPYFYLDTGNTPSQERVELCNRPNGSGDPLFLSLKAGGVGINLTGADT 984

Query: 952 VVLDLWNNPAVEQQAISRHRILQKNNVEYRLITRSTIEKILEMQETKKILVITVTL 1011  
 V+L DLWNNPAVE QA RA+R+QK V+V +L+ CTIEK+ E+QE+KKIL+ V+ +  
 55 Sbjct: 985 VVLYDLWNNPAVEQQAADRAYRWQKNVTVQVIKVAHSTIEKMHLEQSKKILAEVIE 1044

Query: 1012 -GNSTHASMVSVDIREIL 1028  
 G E +S+ ++IR+IL  
 60 Sbjct: 1045 PGEEKLSITTEERDIL 1062

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4823> which encodes the amino acid sequence <SEQ ID 4824>. Analysis of this protein sequence reveals the following:

Possible site: 26

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

-1738-

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.3909(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 674/1031 (65%), Positives = 834/1031 (80%), Gaps = 2/1031 (0%)

10   Query: 1   MSRMIPGRIRNQGIELYEQGLVSLISQBGNNLLKAKVGDCQIEYSLVTETSCDFFARK 60  
       M+R+IPGR+RN+GI+LYSQGLVS                   +L+ +V   Q++Y   B+ C CD F K  
       Sbjct: 2   MARLIPGRVRNBSGILKYQLGLVSPDDNKNKILQIVETVQVOYGADDEDITQCDTHMK 61

15   Query: 61   GYCQHAALEHFLKNDPEGKAILSKVQVOBSQCEKTKTSFGSVFLDSLIINEDDTIKY 120  
       YC+H+AA+E+FLKND +GK L ++ Q ++ TRK TSFGS+FLDSL +NEDD++KY  
       Sbjct: 62   HYCGHAAVEYFLKNDQKGKFLKQLTNTQTKTKETTKMTSPGSLFLDSLIAMNEDDSVKY 121

20   Query: 121   QLSAQGBONFYANDIWNLTAKIRRLPDDRSYVIRDIKAFINTVRKEAYYIGKQYFSTLSL 180  
       +LSA G ++P+++D WH+KI RLPPDRSYVIRDIK FL ++KE +YQIGK YFE LS  
       Sbjct: 122   RLSALGSRSPSSDITWNLSIKINRLPDDRSYVIRDIKGFPLQIKLKGKIFYQIKNIFYEQLSW 181

25   Query: 181   IQPDTSQELIEFLMRLLPSHSSKIDLEFILPNQGRHLSLTRGFFEBGVTIMNALENFSF 240  
       +QPD +SQ LIEFLMR L S + K D E I FN RHL L GFFEBG+ ++L +F+F  
       Sbjct: 182   LQPDSSQALIEFLMRLA-SDTDKNGDNINFPNIAHRLHLPSSGFFEBGHIYLSLYDTFF 240

30   Query: 241   ESDPHQFNHLYFTELEGEDNHYQFKVIVHRQSIETLEIKKDLKPLFANSYLFYRDTYVHL 300  
       ++HL+ + LE E LY+PKV VHR+SIEL+I EK+++ LF N YL Y+DTYHL  
       Sbjct: 241   EGSPQTYHLYFVSLSEAEAGLYEFKVEVHRKSIETLQIAENQVYLFNDYLLYQDTYEL 300

35   Query: 301   NLKQEKMYTAIRSLPIBGLAKHIFDLDQDKLAHLDDFKRIGLVDAPRSPSHDFKV 360  
       LKQ KMY AIRSLPIB GLAKHIFDLDQ KLA L DFK+IGLV+AP+SF+I DF+V  
       Sbjct: 301   TLKQKMYQAIRSLPIBGLAKHIFDLDHDAKLAASLDSPKQIGLVEAPKSAIRDFEV 360

40   Query: 361   NFEFDINSQNSILLQMVFDYGNDLTVHNRQLEQLTFASHKHEEKVFKLEKYGFAPHF 420  
       F+PD+ ++EI Q+FYDGN V ++ LE L FASH K EKK+ +L +GF+P  
       Sbjct: 361   TFQFDLNRDEISQLMFDYGN-VQVSDKASLEALFPASHLKKEEKINRSLAPGFSF 419

45   Query: 421   STSHFAYSAGELYDYFTYMLPQPKMGTVSLAKLESYRLIERPQIDIEAKGSLDLSFD 480  
       +   SA+ELY F+ +P F++G V+LS +++ ++ E P+I I LDISFD  
       Sbjct: 420   YSKKRLTSAKELYTFEETVCFPERLGNVALSTAIQALQVKNPKIAIRRNQGLLISFD 479

50   Query: 481   FSDLLENDVDQALVALFDRNPFYVFNKSGQLVIFDEETKKSATLQGLBARRAKNGHIELD 540  
       FS +END+DQA+ ALF NNPFYV+++GQLV+FD+ET+KVS +LQ LRAR+ KNGH+LQ  
       Sbjct: 480   FSTTIENDIDQAVTALPQNNPFYVSGTQGLVVFDEETQVKSLSQLRLARGLKNGHILD 539

55   Query: 541   NIAAFQISELFANQDNWSPSQHFYQLIEDLRHPKFKFIPOLSVASLRDYLQITGVYRLM 600  
       I A Q+S+LF +V FS+ +L L+HPE F I L V A +RDYQ GV+RLM  
       Sbjct: 540   GIRALQVSKLPEGNTSVHFSLELELAYLHQPETFSIKELFVKAQMDYQRNGVQVLM 599

60   Query: 601   LHNYGFGAILADDWGLKTLQTSPLSTKLTDRSVRLILAPESLILYNNQDEPHKFAPDVD 660  
       L+HYGF GILADDWGLKTLAT++FL++ L DS+VLILSPSSLIYNN DE KF P +D  
       Sbjct: 600   LHNYGFGAILADDWGLKTLQTLAFLASHLSKSDSLILSPSSLIYNNFDECCQKTPQLD 659

65   Query: 661   VAVAYGSKIRKREDEIAERHQVITTSYSSFRQDPETYSRGNVYLILDEAQVMNQATKIA 720  
       V +VYG K RD+II E HQ+ ITSYSFRQDPETTY +YDYLILDEAQV+KNQQTKI+  
       Sbjct: 660   VVVSYGLKQRDQITIEBHQITITSYSSFRQDPETTYQAFHYHILILDEAQVIRNQTAKI 719

70   Query: 721   HSLSRSEVKNCFALSGTPIENKLEIWSIFQIILPGLLQKKKEPLKNPKQVARIKIPV 780  
       H LR+P NCFALSGTPIENK+LEIWSIFQI+LPGLL KKEFLK +QV+RYIKIPV  
       Sbjct: 720   HCLRAFNTANCFALSGTPIENKLEIWSIFQIVLPGLLPDKKEFLKATARQVRYIKIPV 779

75   Query: 781   MRRKKEVLPFLDLIEIMYFNHMTDSQVITYLAQLRQIESIONSSADIANRRKIETLS 840  
       MRR+KE+VLPFLDLIE+HY NEMTD QK IYLAQLR+Q+ I+SSD D++R+KIETLS  
       Sbjct: 780   MRRKKEVLPFLDLIEIYNEMTDBQKATYLAQLRQMDQITNSSDVDISROKIEITLS 839

80   Query: 841   GITRLRQICITFRLPMVDYKESGKLESTROLLTQIKENGHRALISQPRGMLDIAERPMV 900  
       GITRLRQICITP LPMVDY G+SGKLS+SLR LITQIKENGHRALISQPRGMLDIA+EM

-1739-

Sbjct: 840 GITRLRQICDTPSLFMDYQKSKLDSLRILLTQIKENGRHALIFSQPRGMLDLAKQEMT 899  
 Query: 901 AMGLITTYKITGSTPANERHEMTAFNAGSKDAFLISLAKAGVGCLNLTGADTVVILDLWN 960  
 A+GLT+Y+TGSTPANER EMTRAFN GSKDAFLISLAKAGVG+NLTGADTV+LIDLWN  
 5 Sbjct: 900 ALGLTSYQMTGSTPANERQEMTRAFNNGSKDAFLISLAKAGVGCLNLTGADTVVILDLWN 959  
 Query: 961 PAVEMQAISRAHRLGQKENVEVYRLITRGTTIEKKILEMQETKKHLVTVVLXGNETHASMS 1020  
 PAVEMQAISRA+R+GQKENVEVYRLITRGTTIEKKILE+QK+K++LVTVVLXGNE+ ASMS  
 10 Sbjct: 960 PAVEMQAISRAYRIGQKENVEVYRLITRGTTIEKKILEQSKRNLVTVVLXGNESRASMS 1019  
 Query: 1021 VDDIREILGVS 1031  
 ++I+EILG++  
 Sbjct: 1020 IEBIKSILGLN 1030

15 SEQ ID 4822 (GBS369) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 5; MW 120kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 6; MW 142kDa).

The GBS369-GST fusion product was purified (Figure 215, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 303), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1562

25 A DNA sequence (GBSx1654) was identified in *S. agalactiae* <SEQ ID 4825> which encodes the amino acid sequence <SEQ ID 4826>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 30 bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 There is also homology to SEQ ID 1034:

Identities = 34/38 (89%), Positives = 37/38 (96%)  
 Query: 1 MKKEARQIIDLKRNLFKIDVRAQKDEKVFMTACQFS 38  
 +EKEARQ+IDLKRNLFKIDVRAQKDEKVFMTAC+ S  
 40 Sbjct: 1 LKKEARQMIDLKRNLFKIDVRAQKDEKVFMTACQFS 38

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1563

45 A DNA sequence (GBSx1656) was identified in *S. agalactiae* <SEQ ID 4827> which encodes the amino acid sequence <SEQ ID 4828>. This protein is predicted to be phosphoglycerate dehydrogenase (cra2). Analysis of this protein sequence reveals the following:

possible site: 31  
 >>> Seems to have no N-terminal signal sequence

50

-1740-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3709 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GF:BAAB8823 GB:AB016077 phosphoglycerate dehydrogenase  
 [Streptococcus mutans]  
 Identities = 377/436 (86%), Positives = 414/436 (94%)

10

Query: 1 MVLPTVAIVGRPNVGSSTLFNRIAGERISIVEDVSGVTRDRIYTTGEWLNKPSLIDTGG 60  
 M LPTVAIVGRPNVGS LFNRIAGERISIVEDVSGVTRDRIYTTGEWLNK+PS+IDTGG  
 Sbjct: 1 MALPTVAIVGRPNVGSALFNRIAGERISIVEDVSGVTRDRIYTKAEWLNKQFSLIDTGG 60

15

Query: 61 IDDVADPFMEQIKHQADIAMTEADVIVFVSGKEGVTDADEYVSRILYKTNKPVILAVNK 120  
 IDDVADPFMEQIKHQADIAMTEADVIVFVVS KEG+TDADEYV++ILY+T+KPVILAVNK  
 Sbjct: 61 IDDVADPFMEQIKHQADIAMTEADVIVFVVSKEGVTDADEYVAKILYKTHKPVILAVNK 120

20

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILOAIVENLPVEENENPDIIIRFSLI 180  
 VDNPEMR+ IYDFY+LGLGDPY+SS HGIGTGDI+LOAIV+NLP E + E+ DII+FSLI  
 Sbjct: 121 VDNPEMRSAIYDFYALGLGDPYPVSSAHGIGTGDILOAIVNLPAREAQESSDIIRFSLI 180

25

Query: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTNPVDSQGGQYTMIDTAGMRKSGIVY 240  
 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDT F D +QGE+TMIDTAGMRKSGIVY  
 Sbjct: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTTTFDEGGQFTMIDTAGMRKSGIVY 240

30

Query: 241 ENTEKYSVMRRAIDRSDVLMVINAEGIREYDKRIAGFAHEGKGI++VVNKMID I+  
 ENTEKYSVMR+RAIDRSD+VLMV+NAEGIREYDKRIAGFAHE GKG++VVNKMID I+  
 Sbjct: 241 ENTEKYSVMRRAIDRSDVLMVINAEGIREYDKRIAGFAHEAGKGI VVVVNKMIDIAIK 300

35

Query: 301 KDSHTVSQWEADIRDNFQFLSYAPIFVSABTKQRLHKLPEMDIKRISQEQNKRIIPSAVLN 360  
 KDN TV+QWE DIRDNFQ++ YAPI+FVSA TKQRLHKLPEMDIKRISQEQNKRIIPSAVLN  
 Sbjct: 301 KDSHTVAQWETDIRDNFQYIYAPIFVVSATKQRLHKLPEMDIKRISQEQNKRIIPSAVLN 360

40

Query: 361 DVMDAIAINPTPTDKGKRLKIPYATQVAVKPTTFVVFVNEEELMHPSYLRFLNQIRQA 420  
 DV+MDA+AINPTPTDKGKRLKIPYATQV+VKPTTFV+VFNNEELMHPSYLRFLNQIR+A  
 Sbjct: 361 DVMDAIAINPTPTDKGKRLKIPYATQVSVKPTTFVVFVNEEELMHPSYLRFLNQIRQA 420

45

Query: 421 FVFEGTPIILARKKK 436  
 FVFEGTPI LIAKRRK  
 Sbjct: 421 FVFEGTPIILARKKK 436

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4829> which encodes the amino acid sequence <SEQ ID 4830>. Analysis of this protein sequence reveals the following:

45

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3463 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

55

Identities = 403/436 (92%), Positives = 422/436 (96%)

Query: 1 MVLPTVAIVGRPNVGSSTLFNRIAGERISIVEDVSGVTRDRIYTTGEWLNKPSLIDTGG 60  
 MVLPTVAIVGRPNVGSSTLFNRIAGERISIVEDVSGVTRDRIYTTGEWLNK+PSLIDTGG  
 Sbjct: 1 MVLPTVAIVGRPNVGSSTLFNRIAGRRIS:VEDVSGVTRDRIYATGEWLNKQFSLIDTGG 60

60

Query: 61 IDDVADPFMEQIKHQADIAMTEADVIVFVSGKEGVTDADEYVSRILYKTNKPVILAVNK 120  
 IDDVADPFMEQIKHQADIAMTEADVIVFVVSKEGVTDADEYVSRILYKTNKPVILAVNK  
 Sbjct: 61 IDDVADPFMEQIKHQADIAMTEADVIVFVVSKEGVTDADEYVSKILYKTNKPVILAVNK 120

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILOAIVENLPVEENENPDIIIRFSLI 180

-1741-

VDNPEMRNDIYDFYSLGLGDPYP+SSVHGIGTGD+LDAIVENLFVEE EN DIIRPSLI  
 5 Shjct: 121 VDNPEMRNDIYDFYSLGLGDPYP+SSVHGIGTGD+LDAIVENLFVEEAKENDDIIRPSLI 180

Query: 181 GRPIVKGSSLIINAILGSDRVIASPVAGTTRDAIDTFVDSQQEYTMIDTAGMRKSGKVY 240  
 GRPIVKGSSLIINAILGSDRVIASPVAGTTRDAIDTFVDSQQEYTMIDTAGMRKSGK+Y  
 5 Shjct: 181 GRPIVKGSSLIINAILGSDRVIASPVAGTTRDAIDTFVDSQQEYTMIDTAGMRKSGKVY 240

Query: 241 ENTEKYSVMRSMRAIDRSVVLMVINAEGBIREYDKRIAGFAHETGKGIIVVNMWDITIE 300  
 ENTEKYSVMRSMRAIDRSVVLMVINAEGBIREYDKRIAGFAHETGKGIIVVNMWDITIE  
 10 Shjct: 241 ENTEKYSVMRSMRAIDRSVVLMVINAEGBIREYDKRIAGFAHETGKGIIVVNMWDITIE 300

Query: 301 KDNHTVQWSEADIRDQFQLFYAPIIIVSAETKQRLKLPDIKRISEQNKRIPIASVIN 360  
 KDNHTVQWSEADIRDQFQLFYAPIIIVSAETKQRLKLPDIKRISEQNKRIPIASVIN  
 15 Shjct: 301 KDNHTVQWSEADIRDQFQLFYAPIIIVSAETKQRLKLPDIKRISEQNKRIPIASVIN 360

Query: 361 DVIMDAIAINPTTCKGRLKIFATQVAVKPTTFVVFVNEBELMFSYLRFLNQIRRA 420  
 DVIMDAIAINPTTCKGRLKIFATQVAVKPTTFVVFVNEBELMFSYLRFLNQIRRA  
 20 Shjct: 361 DVIMDAIAINPTTCKGRLKIFATQVAVKPTTFVVFVNEBELMFSYLRFLNQIRRA 420

Query: 421 FVFGTPTIILIAIRKK 436  
 F FVFGTPTIILIAIRKK  
 25 Shjct: 421 FVFGTPTIILIAIRKK 436

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 25 vaccines or diagnostics.

### Example 1564

A DNA sequence (GBSx1657) was identified in *S. agalactiae* <SEQ ID 4831> which encodes the amino  
 acid sequence <SEQ ID 4832>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2734 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA00359 GB:AP008220 DnaI [Bacillus subtilis]  
 Identities = 105/313 (33%), Positives = 191/313 (60%), Gaps = 17/313 (5%)

40 Query: 1 MKSVGQALENIGRVP--RNTNDELIQMILADAVAEPIETHQ--LSQREINISMGKFNQF 56  
 M+ +G+L+ P + +++ + + D V F+K++ + Q+I S+K++  
 Shjct: 1 MEPIGRGLQGVTRPDQPKRLGKMKKVMKDQVQAFLENEEVIDQMKTEKSLNLEYE 60

45 Query: 57 LIERQK----PKNKDSYIARQVEPLIUVNBGVIADVSYLE--TRELTAQKQAIISRI 109  
 IE+ K ++++ + +G P LV+N D+ Y E + ++ QKQ +  
 Shjct: 61 -IEQKNCSCYSEDENCENLLBGYHFKLVNNGRSIDIEYECVPKRLDQKQ--QSLM 117

Query: 110 NLVNLPSKRYNRMRTDFDINNESMGKAMSQLLDPVPTPSYNH-KGLYLGDMGVGKSYL 168  
 + + + DI+ SR+ + DP+++Y KGLYLG QVK+K++  
 50 Shjct: 118 KSMYIQQDLGATQQVQDISDPERLAMPQHVDPLESYNETGKNGKLYLYGKPGVQKTFM 177

Query: 169 NAAMARELSERKNGVSTTLHPSFAIDVKNASGTVKDEIDAVKSVPIILDDIGAEQA 228  
 +AA+A EL+E+ S+ +++ P P + +KM++ T+++++ VE+ P+L+LDIGAE  
 55 Shjct: 178 LAALANELAEKE-YSMIVTVPSFVRELKNSLQDQTEELKINMVRTVPMALDDIGAESM 236

Query: 229 TSWVRDEILQVILQHMLRELTPFTSNYSFNLERKGA-NIKGSDHTQAKVMEVRY 287  
 TSWVRDE++ +LQHNN +LPTFF+SN+S +L+ + + +G E +A RME+Y  
 Shjct: 237 TSWVRDEIGTVLQHNSQLPTFSSNPSFDELKIHFTYSGQGEKVEVKAARMERILY 296

60 Query: 288 LAIEFHLEGPMNR 300  
 LA L+G NRR

-1742-

Sbjct: 297 LAAPIRLDGENRR 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4833> which encodes the amino acid sequence <SEQ ID 4834>. Analysis of this protein sequence reveals the following:

5      Possible site: 19  
      >>> Seems to have no N-terminal signal sequence

10      ----- Final Results -----  
                  bacterial cytoplasm --- Certainty=0.1944 (Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 228/300 (76%), Positives = 264/300 (88%)

15      Query: 1    M K S V G A L E N Q G R V P R N T N D E L I Q M I L A D A Q V A E F I K T H O L S Q R E I N I S M S K F N Q F L I E R 60  
                  M + + G +    G +    R    + D + L I Q I L A D + V A F I    H L S Q + I N + S + S K F N Q F L + E R  
                  Sbjct: 1    M E K I G E T M A K L G Q N T R V N S D Q L I Q T I L A D P E V A S F I S Q H L S C E Q I N L S L S K F N Q F L V E R 60

20      Query: 61    Q K F G K D S Q Y I A N G Y E P I L V M E G Y A D V S Y L E T R E L I F A Q K K Q A I S D R I N L V N L P K S Y R N 120  
                  Q K +    K D    Y I A N G Y + P I L    M M E G Y A D V S Y L E T + E L + R A Q K +    A I S + R I    L V + L P K S Y R +  
                  Sbjct: 61    Q K Y Q L K D P S Y I A N G Y Q P I L A M E G Y A D V S Y L E T K E L V E A Q K A I S E R I Q L V S L P K S Y R H 120

25      Query: 121    I R M T D P D I N N E S R M K M S Q L L D F V E T Y P S Y N H K G L Y L Y G D M G V K S Y L A M A M A E L S E R K 180  
                  I    + + D D + N N S R M + A S + L D F V E Y P S    K G L Y L Y G D M G + G K S Y L + A A M A    E L S E + K  
                  Sbjct: 121    I H L S D I D V N A S R M E A F S A I L D F V E Q Y P S A B Q K G L Y L Y G D M G I G S Y L A M M H E L S E K K 180

30      Query: 181    G V S T T L L H F P S F A I D V I A N I S S G T V K D E I D A V K S V P I L I L D D I G A B Q A T S N V R D E I L Q V I 240  
                  G V S T T L L H F P S F A I D V I A N I S + G + V K + E I D A V K + V P + L I L D D I G A B Q A T S N V R D E + L Q V I  
                  Sbjct: 181    G V S T T L L H F P S F A I D V I A N I S N G S V K E E I D A V K N V P V I L I D D I G A B Q A T S N V R D E V I L Q V I 240

35      Query: 241    L Q H R M L E E L P T F F T S N Y S F N D L E R K W A N I K S D E T W Q A K R V M E R V R Y L A E F H L E G P N R 300  
                  L Q + R M L E E L P T F F T S N Y S F    D L E R K W A    I N G S D E T W Q A K R V M E R V R Y L A    E F H L E G N R R  
                  Sbjct: 241    L Q Y R M L E E L P T F F T S N Y S F A D L E R K W A T I K S D E T W Q A K R V M E R V R Y L A E F H L E G A N R R 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1565

40      A DNA sequence (GBSx1658) was identified in *S.agalactiae* <SEQ ID 4835> which encodes the amino acid sequence <SEQ ID 4836>. Analysis of this protein sequence reveals the following:

50      Possible site: 37  
      >>> Seems to have no N-terminal signal sequence

45      ----- Final Results -----  
                  bacterial cytoplasm --- Certainty=0.2660 (Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50      A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4837> which encodes the amino acid sequence <SEQ ID 4838>. Analysis of this protein sequence reveals the following:

55      Possible site: 35  
      >>> Seems to have no N-terminal signal sequence

55      ----- Final Results -----  
                  bacterial cytoplasm --- Certainty=0.2135 (Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>



-1743-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 217/391 (55%), Positives = 309/391 (78%)

5 Query: 1 MMSPIDEFITYIKQNKIVDSNSLIQLYPPIMGSDAMALYDYFVHFDDGIRRHKPSVLN 60  
MM PID FTY+K+K+K+ DS +LIQLYPP+GSDA++Y YP+HFFDDG++RHKPS++LN  
Sbjct: 1 MMKPIDTFTYLRNKKVTLDSVTLIQLYPP+IGSDAVSIYQYPIHFFDDGLQRHKPSDLN 60

10 Query: 61 HLQYCMRPQDALVMLTALDELIVYQATGTYLKLNQAMSNELFLSNFTYRRLLEKIGR 120  
HLQ+GM RF+DAL +LTA+L+VYQ + TYL+ L+Q +S +L F +P Y RLL+ +IGR  
Sbjct: 61 HLQPGMKRFEDALAILTAMELWSVYQLSDTYLTLHLQPLSRDLFFQHPATSRLLLEQKIGR 120

15 Query: 121 VAAVELDMKIPKNARDISKKFTDVSFLGQPKQEVNRKSNVFDLESFKRLMRDGLRPN 180  
VAV+EL + +P AR+ISK+F+D+F G + + FDL SP++LM+RDGI+F +  
Sbjct: 121 VAVSELQVTVPSQARNISKRFSDIPGVQDLTNVQCPQKNFDLSFPQLAVNRDGLQPED 180

20 Query: 181 EKDDVLGIYSVSELYHLNVDYTLQAKQTAINGMLAPQRMKVVQNGQHKIKDQSPFTN 240  
+ D++ +YS++E Y + W+DTYQ+AK TA+NG I P+R+ ++N+ ++P+ E  
Sbjct: 181 NQKDIISLSYLAEGYDMTWFDTYQAKATAVNGKIFERLLAKNGQSMTPSKENPSQAE 240

25 Query: 241 KVILRESKNSALVPLEKIKRSRKAVTTSGEKTLEELAKMNFLEVINVMVLYTNTKTK 300  
++ILRE+K DSAVLPLEKIK++R+A T E+ L+ LAKMNFLE+VINVMVLYT NTKTK  
Sbjct: 241 QILRLKQDSALVPLEKIKARRATITKDERILQLAKMNFLE+VINVMVLYTNTKTK 300

30 Query: 301 SANLNKAYIMKVANDPAQNMVTAREDAVLKIRDPDQKVRTKTETKTKQSNVPWSNPDY 360  
SANTL K+Y++K+ANDFA+Q V TAE+A++ +R F+D++ R +++ K QSNVP+WSNPDY  
Sbjct: 301 SANLQKSYVLQANDPAQKQVSTAEAVLVRAFTIRQSRQSKVTKTSQSNVPKWSNPDY 360

35 Query: 361 KDEVSPKEIELEQPKTDALKRLERLKGDE 391  
++ S E++ +L+QFK ALKRL E LK G+  
Sbjct: 361 QETTSQEEQAKLDQFKQAAKRLNLEKGGD 391

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1566**

A DNA sequence (GBSx1659) was identified in *S. agalactiae* <SEQ ID 4839> which encodes the amino acid sequence <SEQ ID 4840>. Analysis of this protein sequence reveals the following:

40 Possible site: 19  
>>> Seems to have no N-terminal signal sequence

---- Final Results ----  
bacterial cytoplasm --- Certainty=0.4485 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06865 GB:AP001517 unknown conserved protein [Bacillus halodurans]  
Identities = 80/150 (53%), Positives = 115/150 (76%)

50 Query: 1 MRCPKCGYNKSSVDSRQAEQOTTIRRRKECEKQGNFTTFRLEKPLPLAVI KKGITREQ 60  
MRCP C +N + V+DSR A EG +IRRRRECE C +RFTTPE +EE+PL+V+KKGITR++  
Sbjct: 1 MRCPACHINGTRVLDSPRAIEGRSIRRRKECESCNIRFTTFRMIEEPLIVVKIGSTRQR 60

55 Query: 61 PSRDKILNGIIQSAQKRPVSSIEDIRNCILRIERKIRSEYREDEVSSITIGNLVMDLAEILD 120  
PS DKLL G+I++ +KRFV E +E + +ER++R + ++EV S IG LVH+ LA +D  
Sbjct: 61 PSSDKILNGLIRACEKRPVLETLRGIVNEVERELRGQKNEVDSKRIEGLVMERLANVD 120

60 Query: 121 EITYVRFASVYKSPKOVDEIEELQQITGR 150  
++ YVRFASVY+ FKD++ + L+++ +R  
Sbjct: 121 DVAYVRFASVYQPKDINVFQELKILMER 150

-1744-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4841> which encodes the amino acid sequence <SEQ ID 4842>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4365(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 131/155 (84%), Positives = 143/155 (91%)

15  Query: 1  MRCFKGYNKSSVVDGRQAEGTTIRRRRECEKGNRPTTFERLEELPLLVINKDGTREQ 60
      Sbjct: 1  VRCFKCNYHKS SVVDGRQAEDGNTIRRRRECEQCHTRPTTFERVEELPLLVINKDGTREQ 60

20  Query: 61  FSRDKILNGIIQSAQKRPVSSEDIENCILRIERKIRSEYEDEVSSITIGNLVMDLAELD 120
      Sbjct: 61  FSRDKILNG++QSAQKRPVSS DIEN I RIE++R+ YE+EVSS IGNLVMDLAELD 120

      Query: 121 EITYVRFASVYKSPKVDIEELLQQITRVRGKK 155
      Sbjct: 121 EITYVRFASVYKSPKVDIEELLQQIT RVR KK 155

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1567

30 A DNA sequence (GBSx1660) was identified in *S.agalactiae* <SEQ ID 4843> which encodes the amino acid sequence <SEQ ID 4844>. This protein is predicted to be CsrS (mtrB). Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence

35  INTEGRAL  Likelihood = -11.30  Transmembrane  22 - 38 ( 18 - 43)
      INTEGRAL  Likelihood = -9.66  Transmembrane  189 - 205 ( 187 - 212)

----- Final Results -----
      bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2109> which encodes the amino acid sequence <SEQ ID 2110>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.

50  INTEGRAL  Likelihood = -6.32  Transmembrane  196 - 212 ( 189 - 214)

----- Final Results -----
      bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

55  Identities = 248/501 (49%), Positives = 363/501 (71%), Gaps = 4/501 (0%)

```

-1745-

Query: 1 MNKKQXQFIGVKQPLSKKLSQVLVFFSLPIVFSVLVYTSATRYVLHREKINVGSRLEK 60  
 M+N+K + K L K+Ls + P+LFF +P+ P+++ Y+S ++L +EK +V +++  
 Sbjct: 1 MENQKQKQKQKYNKSLPKRLNIFVLPFCIPSAPTLIAYSSINYPLLKKERQSVFQAVNI 60

5 Query: 61 TRVLSQANSSLTSDILEILYNQVADDIYPHQRNGIVTGESIDISILYNQMTLYD 120  
 RVRLG+ +S+ T +++ E+LY ++ + ++R+ I + L NQ++ +Y+  
 Sbjct: 61 VVRVLSVDNFTLENLAELVLYKNDKTHLRIDDRKGSRVISERDITNTLDANQIYVYN 120

10 Query: 121 VNRKPFVST-LRTGMPITIGSKNGKVIISKVADM-EGFVGTKALYSQKTOQLLGVQVPIPN 178  
 ++++ +P+T P + +G+V +D GF T+ +YS +TG+ +GVVQ+P++  
 Sbjct: 121 IDKQVPTTDMRESSGLKGIKGVYHDHIEDQYKGFSPMTQVYSNRTGKPFVGVVQVFD 180

15 Query: 179 LGRYYSMRQNIIVFLIMMEVLGTVALVINSATKRIVRVKNLHDLMHQISENPNLEI 238  
 LG YY +R ++ +L+++E+ GT LA ++I T+R ++P+ NLH++M ISENPN+L +  
 Sbjct: 181 LGRYVIRARLLFWLVLVVELFQTSLAYLIITLITRFLKPLNHLHEVMKNISENPNLN 240

20 Query: 239 RSKVRSDEIGELSRIFDGMLOQLEDYTRRQSQFTSDVSHKLRTPVAVVGKHIGLQRWG 298  
 RS + S DEI ELS IFD MLD+IE +T+ QS+FTSDVSHKLRTPVA++KGHIGLQRWG  
 Sbjct: 241 RSDISGDEIEELSVIFDNLKLEHTHTKLQSRFTSDVSHKLRTPVAIIGKHIGLQRWG 300

25 Query: 299 KDDPEILRESLAAAYHEADRMISIMINDMLAMIRVQGSLELQOEVTDLSSISVVIENFR 358  
 KDD +ILEESL A HEADRM++MINDML+MIRVQGS E HQ+++T L SI V+ NFR  
 Sbjct: 301 KDDSDILEESLTATAHEADRMATMINDMLCMIRVQGSFEGHQNMTVLEDSIETVVGNGR 360

30 Query: 359 ILREDFQIFERNISDIWNGKIYKIHFEQALMILIDNAIKYSPSYKEVSVLVDNDPAT 418  
 +LEDF F +++ + +YK HFEQALMILIDNA+KYS K++++ LSV  
 Sbjct: 361 VLREDFPIIQSENPKTI-ARIYKNHFEQALMILIDNAIKYSRKEKKIAINLSVTQKEA 419

35 Query: 419 VV-VDKGEGISDEIDIEFIDRFYRTDKSRNRESTQAGLIGLGSVFKQIMDAYHLKVDIK 477  
 +V +V+DKGEGIS EDIE IF+RFYRTDKSRN STQAGLIGLGS+ KQI+D YHL++ ++  
 Sbjct: 420 IVRVQKGEIGISDEIDIEHIFRFYRTDKSRNRTSTQAGLIGLGSILKQIVDGYHLQMKVE 479

Query: 478 SELNQOTSEFIVRIPIKGFET 498  
 SELN+G+ FI+ IP+ + +E+  
 Sbjct: 480 SELNKGSVFILHIPLAQSKES 500

A related GBS gene <SEQ ID 8845> and protein <SEQ ID 8846> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 5  
 SRCLG: 0  
 MoG: Length of UR: 5  
 Peak Value of UR: 0.74  
 Net Charge of CR: 2  
 MoG: Discrim Score: -10.19  
 45 GvH: Signal Score (-7.5): -3.66  
 Possible site: 35  
 >>> Seems to have no N-terminal signal sequence  
 Amino Acid Composition: calculated from 1  
 ALOM program count: 2 value: -11.30 threshold: 0.0  
 50 INTEGRAL Likelihood = -11.30 Transmembrane 22 - 38 ( 18 - 43)  
 INTEGRAL Likelihood = -9.66 Transmembrane 189 - 205 ( 187 - 212)  
 PERIPHERAL Likelihood = 2.86 405  
 modified ALOM score: 2.76  
 icml HYPID: 7 CFP: 0.552  
 55 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5522 (Affirmative) < succs  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succs  
 60 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succs

-1746-

SEQ ID 8846 (GBS321) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 6; MW 84kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 2; MW 58.7kDa).

GBS321-GST was purified as shown in Figure 220, lane 3.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1568

- A DNA sequence (GBSx1661) was identified in *S. agalactiae* <SEQ ID 4845> which encodes the amino acid sequence <SEQ ID 4846>. This protein is predicted to be CsrR (trcR). Analysis of this protein  
10 sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2649 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

- A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3259> which encodes the amino acid  
20 sequence <SEQ ID 3260>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3226 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

- ```
Identities = 193/229 (84%), Positives = 211/229 (91%), Gaps = 1/229 (0%)

Query: 1  MGKILIIIEDEKNLARFVSLLEHGYDVVETNGREGLDTALEKDFDLILLDLMLPEMD 60
Sbjct: 1  MKKILIIIEDEKNLARFVSLLEHGYDVVETNGREGLDTALEKDFDLILLDLMLPEMD 60

Query: 61  GFETTRRLQAETKTYIMMTARDSVMDIVAGLORGADDYIVKPFATIELLARVAIRFRQ 120
Sbjct: 61  GFETTRRLQAETKTYIMMTARDSVMDIVAGLORGADDYIVKPFATIELLARVAIRFRQ 120

Query: 121  EITKTKENGDSGSRDLSTNHNRSAMRGDSIELTKREFDLLNVLMTNMRVMTREEL 180
Sbjct: 121  DTRSE-KKVPSSQGIYRDLVLNPNQSRVNRGDSIELTKREFDLLNVLMTNMRVMTREEL 179

Query: 181  LEHVNYKDVAASTHVVDVYIRYLRGKIDIPGRSYYIQTVRGMGVVIREK 229
Sbjct: 160  LSNVNYKDEAVSTHVVDVYIRYLRGKIDIPGRSYYIQTVRGMGVVIREK 228
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1569

A DNA sequence (GBSx1662) was identified in *S. agalactiae* <SEQ ID 4847> which encodes the amino acid sequence <SEQ ID 4848>. Analysis of this protein sequence reveals the following:

-1747-

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5                   bacterial cytoplasm --- Certainty=0.3864 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:AAG32547 GB:U12643 YlbN-like hypothetical protein [Streptococcus gordonii]  
 Identities = 91/174 (52%), Positives = 133/174 (76%), Gaps = 3/174 (1%)

- Query: 3 LITEIKKSPGLYFDKKIDIKESIMERNHSEIMDISIQVSGHVYEDGLYLDDYNNAYDIT 62  
 + EI+K+P+GL P+KK+D+ E L ER++EI+D+ DI SG YEDGLY LDY ++Y IT  
 15 Sbjct: 4 IQEIRKNPDCIAFEKKKLDAELKERNAILDVQDIVASGRAQYEDGLYFLDYELSTYIT 63
- Query: 63 LPSRSRMKPFVVLSEKQTINSVFIEARNVSTKKELVDQELVLILEEDDINLEESVIDNILL 122  
 L SRSRM+PV E +NE+P+E V++ +E++DQ+LVL +E +IN+ ESVDNILL  
 20 Sbjct: 64 LASSRSMEFVERKESYL/NSEIPMEDGQVAS-QEMIDQDLVLITENGEINVAESVDNILL 122
- Query: 123 NIPLRLVL-AADEVGVADLSGKNWSLMTKEQYEEKQAKEKEKSNPFAALBGMFD 175  
 NIPL+VL AA+E G + +G+W+MTE Y++ QA++KB++PFA L+G+PD  
 Sbjct: 123 NIPLKVLPAEEBAGSDLP-TGRDWQVNTEDDYQKYQAEKKEKSNPFAALBGMFD 175

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4849> which encodes the amino acid sequence <SEQ ID 4850>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 30                   bacterial cytoplasm --- Certainty=0.3032 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 An alignment of the GAS and GBS proteins is shown below.

Identities = 85/175 (49%), Positives = 135/175 (77%)

- Query: 1 MLITEIKKEPEGLYFDKKIDIKESIMERNHSEIMDISIQVSGHVYEDGLYLDDYNNAYD 60  
 + ++EI+K +P+GL PD+ D+K L+ER +I+DI ++ G+V Y+ GLYLDDY ++Y+  
 40 Sbjct: 3 LAISEIRKHPDGLSPDLICDVKEMLLERDQIIDHAKVAVGNVRYDKGLYLDDYQLSYS 62
- Query: 61 TLPSSRSRMKPFVVLSEKQTINEVFIEARNVSTKKELVDQELVLILEEDDINLEESVIDNI 120  
 + LPSSRSRM FV LSE Q I E+FIRA +++ KKELV+ LVL+L++D INLEES++DNI  
 45 Sbjct: 63 VILPSSRSRMVPCLESVGHQELFIRATLDADKKELVEDNLVLVDKDAINLEESVIDNI 122
- Query: 121 LLNIPLRLVLADEVGVADLSGKNWSLMTKEQYEEKQAKEKEKSNPFAALBGMFD 175  
 LL IP++VL +E + +G+NW+++TE+ Y+ +++++NPFAL+G+PD  
 Sbjct: 123 LLAIPVQVLTEEEKSKKELBAGQNMVAVLTEEDYQCLKEEKQKNNPFAALBGMFD 177

- 50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1570

A DNA sequence (GBSx1663) was identified in *S.agalactiae* <SEQ ID 4851> which encodes the amino acid sequence <SEQ ID 4852>. This protein is predicted to be heat shock protein (hspX). Analysis of this protein sequence reveals the following:

possible site: 25  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood =-11.30 Transmembrane 195 - 211 ( 190 - 221)

-1748-

INTEGRAL Likelihood = -11.09 Transmembrane 43 - 59 ( 31 - 62)  
 INTEGRAL Likelihood = -3.61 Transmembrane 153 - 169 ( 153 - 174)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5522 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAB70525 GB:AF017421 putative heat shock protein HtpX  
 [Streptococcus gordonii]  
 Identities = 220/297 (74%), Positives = 261/297 (87%), Gaps = 1/297 (0%)

15 Query: 1 MLVQQIASNKRKTIVLLVFFCLLAIGAAGVYLGSYQPLVLAIIIGVTVNMFQ 60  
 ML++QIA+NRK+T LL+ FF LLA IGAA GYL + S G++A IIG+IYA++MIFQ  
 Sbjct: 1 MLFEQIAANKRRTWFLVAFVALLAIGAAAGVYLMNSFLOGVILAFIIGLYALTMTFQ 60

20 Query: 61 STNVVSMNNAREVTEDEAPYFHVIEDNMAIAQIIMPRVFIIVEDDSINAFATGSKPEN 120  
 ST VVSMNN AR+V+S EAP +HIV+DNAM+AQIPMRV+IVEDDS NAFATGS PEN 120  
 Sbjct: 61 STEVVSMNGARQVSQSEAPELYHIVQDNAMVAQIMPRVFIIVEDDSNPAFATGSKPEN 120

25 Query: 121 AVAATTGLLAVMNRRELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRMLFYG 180  
 AVAATTGLL +MNRRELEGVIGHEVSHIRNYDIRISTIAVALASA+T+ISS+ RM++YG  
 Sbjct: 121 AVAATTGLLRMNRRELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMFYG 180

30 Query: 181 GGRRRDDDDREGG-NILVLI FSI LSLIAPLAASVLQLAISQREYLDASSVELTRNPQ 239  
 GGRRR+D +D G +L+L+FS+++ ILAPLAA+LWQLAISQRE+LDASSVELTRNPQ  
 Sbjct: 181 GGRRRDDDDSGGLSLMLVFSLLIALIAPLAATLVQLAISQREYLDASSVELTRNPQ 240

35 Query: 240 GMISALEKLDRESEPMGHFVDDASAALYINDPTKGLKSLFYTHPPIDRIERLEHM 296  
 GMI AL+KLD SEP VDDASAAALY+DP KK GL+ LFYTHPFI++R+ERL M  
 Sbjct: 241 GMIRALQLKLDNSEPMHRHVDDASAAALYISDPKKEGGLKLFYTHPPISERVERLRM 297

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4853> which encodes the amino acid  
 35 sequence <SEQ ID 4854>. Analysis of this protein sequence reveals the following:

Possible site: 31

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

40 INTEGRAL Likelihood = -9.77 Transmembrane 197 - 213 ( 192 - 223)  
 INTEGRAL Likelihood = -8.33 Transmembrane 43 - 59 ( 33 - 61)  
 INTEGRAL Likelihood = -3.82 Transmembrane 153 - 169 ( 153 - 174)

----- Final Results -----

45 bacterial membrane --- Certainty=0.4906 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB70525 GB:AF017421 putative heat shock protein HtpX [Streptococcus gordonii]  
 Identities = 206/298 (69%), Positives = 257/298 (85%), Gaps = 1/298 (0%)

50 Query: 1 MLVQQISQNKQRITVLLVGFPAALALIGASAGYLLDNYANGVILVIGVITYATSMIFQ 60  
 ML++QI+ NK+RT LLV FPALALALIGA+AGYL +++ G++A +IG+IYA +MIFQ  
 Sbjct: 1 MLFEQIAANKRRTWFLVAFVALLALIGAAAGVYLMNSFLOGVILAFIIGLYALTMTFQ 60

55 Query: 61 STSLVSMNNAREVTEKEAGPFHVIEDNMAVAQIIMPRVFIIVEDDSINAFATGSSPQNA 120  
 ST +VMEMN AR+V+E+EAP +HIV+DNAMVAQIIMPRV+I+ED S NAFATGS+P+NA 120  
 Sbjct: 61 STEVVSMNGARQVSQSEAPELYHIVQDNAMVAQIIMPRVFIIVEDDSNPAFATGSKPEN 120

60 Query: 121 AVAATTGLLEVMNRRELEGVIGHEVSHIRNYDIRISTIAVALASAVTVISSIGSRMLFYG 180  
 AVAATTGLL +MNRRELEGVIGHEVSHIRNYDIRISTIAVALASA+T+ISS+ GRM+MYG  
 Sbjct: 121 AVAATTGLLRMNRRELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMFYG 180

Query: 181 GGRRRQDDGDDVLRITITLLSLSLIAPLASVLQLAISQREYLDASSVELTRNPQ 240  
 GG RR+ D DD L ++ L+ SL+++LAPL A+L+QLAISQRE+LDASSVELTRNPQ

-1749-

Sbjct: 181 GG--RRRDRDDGGSLGLMLVFSLLAIILAPLAATVQLAISRQREYLDASSVELTRNP 239  
 Query: 241 QGMIALEKILQLQSPMKHFVDDASALYINPRKRSFSSLSFTHPTTIERIERLKN 298  
 QGMIALEKILQLQSPMKHFVDDASALYINPRKRSFSSLSFTHPTTIERIERLKN 298  
 Sbjct: 240 QGMIRALQKLEINSEPMIRIVDDASALYISDPKCKGGLQKLFYTHPTTIERIERLKN 297

An alignment of the GAS and GBS proteins is shown below.

Identities = 233/298 (78%), Positives = 262/298 (87%), Gaps = 2/298 (0%)

Query: 1 MLYQQIAENKRRKTVLLIVFFCLAAIGAANGVYLVIGSYQFGLVALLIGVIYAVSMIFQ 60  
 MLYQQI+ NK++TVVLL+ FF LLA IGA+ GYL+L +Y GIALAL+IGVIYA SMIFQ  
 Sbjct: 1 MLYQQISQNKQRTVLLVGSFFALLALIGASAGYLLLDNYMGLVALLIGVIYATSMIFQ 60  
 Query: 61 STNVVMNNAREVTEDEANFYPHIVDMAMIAQIMPRVFIVEDOSLNAFATGSKPEN 120  
 ST++VMNNAREVTE EAP +PHIVEDMAM+AQIMPRVFI+ED SINAFATGS P+NA  
 Sbjct: 61 STSLVMNNAREVTEKEAPGPFPHIVDMAMVQIMPRVFIIEDSINAFATGSSPCNA 120  
 Query: 121 AVAANTGLLAVNRREELEGVIGREVSHIRNYDIRISTIAVALASAVTLISSIGRMLFYG 180  
 AVAANTGLL VNRREELEGVIGHE+SHIRNYDIRISTIAVALASAVT+ISSIG RML+YG  
 Sbjct: 121 AVAANTGLLEVNRREELEGVIGHEISHIRNYDIRISTIAVALASAVTLISSIGRMLFYG 180  
 Query: 181 GG--RRRDRDDGGNIIIVLPSIILSLIAPLAASIVQLAISRQREYLDASSVELTRNP 238  
 GG R+RDD +D I+ L+ S+LSL+LAPL ASL+QLAISRQREYLDASSVELTRNP  
 Sbjct: 181 GGSRRDRDDGGDVLRIITLLLSLLSLLAPLAVASLIQLAISRQREYLDASSVELTRNP 240  
 Query: 239 QGMIALEKILQSPMKHFVDDASALYINDPTRKCKGLSLFYTHPTTIERIERLKN 296  
 QGMIALEKILQSPMKHFVDDASALYINDPTRKCKGLSLFYTHPTTIERIERLKN 296  
 Sbjct: 241 QGMIALEKILQSPMKHFVDDASALYINPRKRSFSSLSFTHPTTIERIERLKN 298

A related GBS gene <SEQ ID 8847> and protein <SEQ ID 8848> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Cterm: 10  
 MoG: Discrim Score: 9.61  
 GvR: Signal Score (-7.5): -0.97  
 Possible site: 25  
 >>> Seems to have a cleavable N-term signal seq.  
 ALON program count: 3 value: -11.30 threshold: 0.0  
 INTEGRAL Likelihood = -11.30 Transmembrane 195 - 211 ( 190 - 221)  
 INTEGRAL Likelihood = -11.09 Transmembrane 43 - 59 ( 31 - 62)  
 INTEGRAL Likelihood = -3.61 Transmembrane 153 - 169 ( 153 - 174)  
 PERIPHERAL Likelihood = 5.89 87  
 modified ALON score: 2.76  
 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5522 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

73.8/88.3% over 296aa

imported

SP|030795| PUTATIVE HEAT SHOCK PROTEIN HTPX. Insert characterized  
 GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus  
 gordonii} Insert characterized  
 PIR|T48855|T48855 probable heat shock protein HtpX - Streptococcus gordonii Insert  
 characterized

ORF02338(301 - 1188 of 1488)  
 SP|030795|HTPX\_STRG(1 - 297 of 297) PUTATIVE HEAT SHOCK PROTEIN  
 HTPX.GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus  
 gordonii}|PIR|T48855|T48855 probable heat shock protein HtpX [imported] - Streptococcus  
 gordonii





-1751-

YE TLEK+ +LRAQVA A++PA+AM AS+ LTRQ+ S I AVAE+YPOLKAN +++KLQ  
 Sbjct: 61 KYBQATLEKVTQLRAQVASASSPADAMKASDALTRQISGIFVAESYDOLKANENYKLQ 120

Query: 121 BELTNTENKISYSRQLYNTTTSYNNVKLETPPSNIVGKLFQPKPSQFLSTPEEEKVQPKV 180  
 BELTNTENKISYSRQLYN+ NNNVKL+ PPSN++ +F P+P+ FL TPEEEK VPVK  
 Sbjct: 121 BELTNTENKISYSRQLYNSVAGNYNVKLQAFPSNVIAGMFAPRPADFLSTPEEEKAVPKV 180

Query: 181 SF 182  
 F  
 Sbjct: 181 DF 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4857> which encodes the amino acid sequence <SEQ ID 4858>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have a cleavable N-term signal seq.  
 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA04350 GB:U66186 Lema [Listeria monocytogenes]  
 Identities = 91/181 (50%), Positives = 121/181 (66%), Gaps = 2/181 (1%)

Query: 5 LILVVLGVLAALMIMISYNSLVKSRMHTKEANSQIDVQLKRRNDLIPNLIVETVGYAYSE 64  
 +I + V+ +L L YNSLVK R E W+QIDVQLKRR DLIPNL+ETVGYA +E  
 Sbjct: 5 IIALAVVILVLIYPLGYNLSVKYRNVDVETWQIDVQLKRRFDLIPNLIVETVGYAKHE 64

Query: 65 QKTFEKITDLARVAN--ASTPQETMAASNELSKQVTSLEFAVBNYDOLKANENFLKQSE 122  
 ++T ++ + R ++ A Q + A N LS + S+FA+ E YPOLKAN +P++LQ  
 Sbjct: 65 KETLTVQVIEARNKMEVPADNRQGGTEADNMLSGALKSI FALGEAYDOLKANTSFIELQH 124

Query: 123 ELTNTENKISYSRQLYNSTTSYNNVKLETPPSNIVGKLFQPKPSQFLSTPEEEKVQPKV 183  
 ELT TENK++YSRQLYN+T YN +++S P+NI KL F + L PE B+ PKVEF  
 Sbjct: 125 ELTNTENKVAYSRLYNITVTMTYNTKVSQVPTNIVAKLENFTDRDMLSPFEVERVAPKVEF 185

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/181 (74%), Positives = 165/181 (90%)

Query: 4 MILIAIALFVWILVAYNSLVSRMHTKESNSQIDVQLKRRNDLIPNLIVETVGYAYSE 63  
 +I++ ++ + +WL+++YNSLV+SRMHTKE+NSQIDVQLKRRNDLIPNLIVETVGYA+YE  
 Sbjct: 5 LILVVLGVLAALMIMISYNSLVKSRMHTKEANSQIDVQLKRRNDLIPNLIVETVGYAYSE 64

Query: 64 GKTLKXIAELRAQVAKANTPAEAMTASNELTRQSLAVAEVBNYDOLKANNSFVKLQSE 123  
 KT EKI +LRA+VA A+TP E M ASNEL++Q++S+ AVAENYDOLKAN +P+KLQSE  
 Sbjct: 65 KQTFEKITDLARVANASTPQETMAASNELSKQVTSLEFAVBNYDOLKANENFLKQSE 124

Query: 124 TTNENKISYSRQLYNITTSYNNVKLETPPSNIVGKLFQPKPSQFLSTPEEEKVQPKVSF 184  
 TTNENKISYSRQLYN+TTSYNN++L+PSPNI GKLRGKPS+FL+TPE EKEVPK F+P  
 Sbjct: 125 TTNENKISYSRQLYNSTTSYNNVKLETPPSNIVGKLFQPKPSQFLSTPEEEKVQPKVSF 185

A related GBS gene <SEQ ID 8849> and protein <SEQ ID 8850> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0  
 MoG: Discrim Score: 14.63  
 GVH: Signal Score (-7.5): -3.19  
 Possible site: 20  
 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 1 value: -15.44 threshold: 0.0  
 INTEGRAL Likelihood = -15.44 Transmembrane 4 - 20 ( 1 - 27)



-1753-

bacterial cytoplasm --- Certainty=0.2430(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 10079> which encodes amino acid sequence <SEQ ID 10080> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>P:GAB16137 GB:Z99124 glucose-inhibited division protein [Bacillus subtilis]  
 Identities = 130/239 (54%), Positives = 170/239 (70%), Gaps = 4/239 (1%)

10 Query: 5 MTPQAFYQVLIHGHITLTDKQKQFSTYFRLIVWNEKINLTAITDKREEVYLKHPYDSIA 64  
 M + F L E G I+L+ +Q +QFE Y+ +LVEWNEKINLT+IT+K+EVYLKHPYDSI  
 Sbjct: 1 MNIREFTSLARKGISLSFQLEQFELYDMLVWNEKINLTISITEKKSVYLKHPYDSIT 60

15 Query: 65 PILQGVYID-NSPLSILDIGAGAGFSPIMKILYPEIDITIIDSLNKRINFLNLANELEL 123  
 Y+D N +I D+GAGAGFSP+P+KI +P + +TI+DSLNRKI FL L+ L+L  
 Sbjct: 61 AAF--VYDFNQVNTICDVGAGAGFSPLEIKICFFHLAVTIVDSLNRKITFLKLSAQL 118

20 Query: 124 SGVHFHGRAEDFGQDRVFRAKFDIVTARAVAKMQLAELTIPFLKVNRLIALKAAAE 163  
 F H RAE FGQ + R +DIVTARAVA++ VL+EL +P +K NG +ALKAA+AE  
 Sbjct: 119 ENITTECHDAETFGQDRVRESYDIVTARAVRLSVLSRLCLPLVKNGFLVALKAAAE 178

Query: 184 EELISAELKALITLFSQVTVNKNYKLP-NGDDRNITIVSKKKEPTNKYPRKAGTPNKKPL 241  
 EEL + +KA+ TL ++ ++LKP DRNI ++ K K TP KYPRK GTPNK P+  
 25 Sbjct: 179 EELNAGKKAITLGGLENIHSFCLPIESDRNIMVIRKINTPTPKYPRKAGTPNKSPI 237

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4861> which encodes the amino acid sequence <SEQ ID 4862>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4862(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/237 (71%), Positives = 202/237 (84%)

40 Query: 5 MTPQAFYQVLIHGHITLTDKQKQFSTYFRLIVWNEKINLTAITDKREEVYLKHPYDSIA 64  
 MTPQ FY+ L E G +L+ KQK+QF+TYF+ LVWNE KINLTAIT+ EVYLKHPYDSIA  
 Sbjct: 1 MTPQDFYRTLESDGFSLSKQKEQFTYFSLVWNEKINLTAITTEENEVYLKHPYDSIA 60

45 Query: 65 PILQGVYIDNSPLSILDIGAGAGFSPIMKILYPEIDITIIDSLNKRINFLNLANELEL 124  
 PILQ++ N P+ +LDTGAGAGFSP+PMKIL+P +++TIIDSLNKR+FL +LA EL L  
 Sbjct: 61 PILQGLFLANEPKILLDIGAGAGFSPIMKILFPNLEVTIIDSLNKRISFLTLAQLGL 120

Query: 125 GVHFFHGRAEDFGQDRVFRAKFDIVTARAVAKMQLAELTIPFLKVNRLIALKAAAE 164  
 VHFHGRAEDFGQD+ FR +FD+VTARAVA+MQVL+ELTIPFLK+ G+LIALKA AA++  
 50 Sbjct: 121 NVHFFHGRAEDFGQDKAPRGQFDVTARAVA+MQVLSRLTIPFLKIGGK+LIALKAQADQ 180

Query: 185 ELISAELKALITLFSQVTVNKNYKLPNGDDRNITIVSKKKEPTNKYPRKAGTPNKKPL 241  
 EL A+ AL LP +V N +Y+LPNGD R ITIV KKKETPNKYPRKAGTPNKKPL  
 Sbjct: 181 ELEEAKNALCLLPKGVKNHSTYQLPNQDSRFITIVKGGKETPNKYPRKAGTPNKKPL 237

55

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1754-

**Example 1573**

A DNA sequence (GBSx1667) was identified in *S.agalactiae* <SEQ ID 4863> which encodes the amino acid sequence <SEQ ID 4864>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1574**

A DNA sequence (GBSx1668) was identified in *S.agalactiae* <SEQ ID 4865> which encodes the amino acid sequence <SEQ ID 4866>. This protein is predicted to be v-type sodium ATP synthase subunit j. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -10.14 Transmembrane 371 - 387 ( 362 - 391)
INTEGRAL Likelihood = -7.48 Transmembrane 200 - 216 ( 190 - 217)
INTEGRAL Likelihood = -4.94 Transmembrane 425 - 441 ( 423 - 446)
INTEGRAL Likelihood = -4.67 Transmembrane 327 - 343 ( 325 - 349)
INTEGRAL Likelihood = -3.77 Transmembrane 81 - 97 ( 81 - 98)
INTEGRAL Likelihood = -2.66 Transmembrane 140 - 156 ( 139 - 157)
INTEGRAL Likelihood = -1.33 Transmembrane 55 - 71 ( 53 - 71)
INTEGRAL Likelihood = -0.27 Transmembrane 247 - 263 ( 247 - 263)
INTEGRAL Likelihood = -0.11 Transmembrane 165 - 181 ( 165 - 181)

----- Final Results -----
bacterial membrane --- Certainty=0.5055 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10081> which encodes amino acid sequence <SEQ ID 10082> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA04279 GB:D17462 Na+ -ATPase subunit J [Enterococcus hirae]
Identities = 170/461 (36%), Positives = 262/461 (55%), Gaps = 28/461 (6%)

Query: 12 KTMVARKLISISFIIVILLSSILLSLPIQYANAPKTHYIDHLPTTVMVCVTGLSVFPI 71
K +S + + F + IL G LL + LP F + TH + ID LFT S VCVTGL+
Sbjct: 10 KRLSPVQLIAGFFILLFQGSLLTLTPFFS-RSGESTHIFDALFTATSAVCVTGLTLINT 68

Query: 72 SKVYNGWQGVIAILLMQTGLSLVLTNLSLYTYLLKKRMSINDQTLQLQSAITYNSDLEKK 131
+ + N GQ + + L+ + GGLG + + L + + K+S + + L+ A+ + K
Sbjct: 69 AERWNSAGQFLIMFLIEIGLGRMMIPILFFALAKKISPSMRIVLKEALINLEMSGVIK 128

Query: 132 YLYMIKVTITLLEVLAAISILAIIDFIPRFLGHGIPNSIFLAVSAFCNAGFDNLENTSLAQ 191
+ I K + + +V+ A L+ + FIP FG GI+ SIP AVS+FCNAGFD L + LA
Sbjct: 129 IMIYILKFAVVVQVIGAVALSUVFIPRFGWAGKIWPSIFHVSFECNAGFDLLGDSLLAD 188

```

-1755-

Query: 192 FKINPLVNIIVCFLLISGLGFVAVWKDLEIATICTSEKGPGLIKTPPKRLSNHSLKVLKT 251  
+ N + ++V LII+GGLGF VM+D+ + E+ K+++ HSK+ L  
Sbjct: 189 -QTNVYLIMVVEALIIAGGLGFIVWRDIL-----SYHR-----VKKTLISKVALSV 234

5 Query: 252 TTTILLVTTLISWLELPGNFRTIANLSLPKQLMVSFPQTVMRTAGSFSTIDYCTQDPATN 311  
T ++L+ G+L +L+ N T+ + ++L +FF +VT RTAG+ +IDY Q A  
Sbjct: 235 TALLLIGGFIL-FLITERNGTLVLKGTFTTERLANTFVMSVPTRTAGYSIDYLMQMSHAGL 293

10 Query: 312 LVYIIQMLIGGAPGGTAGGFKVTIAIIILLKFAELSGSQSVTFHYRTIPSSIIKQTLST 371  
++ + M IGG G TAGG K T + ILL+ A G+++ RTI + + L  
Sbjct: 254 ILMFLMYIGTSGSTAGGLKTTITLIGILLIQMHAMFGKTRARAFGRTIRQAIV---LRA 350

15 Query: 372 LTTFFFI--LISGYLLLELNPHIDPPS----LFRASSALATVGVVMTNTQTLIGRI 425  
LT FF+ L ++L+ I S + FE SA TVG+IM T LTL G++  
Sbjct: 351 LTLFFVTLSCVAINVLSVTETIPTKSGIETAFVFSAPGTVGLIMGLTDLTLIGKL 410

Query: 426 VIMFLMFIGRVGPITVLLSILQK--KKEKTHVAETIILG 463  
VI+ IM+IGRVG +TV+LS+L K E Y R I+LG  
Sbjct: 411 VIISLMYIGRVGIMTVLLSLVKNRANRYKYPEESIMLG 451

20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4867> which encodes the amino acid sequence <SEQ ID 4868>. Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -15.12 Transmembrane 371 - 387 ( 364 - 396)  
INTEGRAL Likelihood = -7.32 Transmembrane 20 - 36 ( 18 - 42)  
INTEGRAL Likelihood = -6.53 Transmembrane 425 - 441 ( 417 - 446)  
INTEGRAL Likelihood = -6.16 Transmembrane 89 - 105 ( 81 - 106)  
INTEGRAL Likelihood = -5.79 Transmembrane 200 - 216 ( 196 - 223)  
30 INTEGRAL Likelihood = -3.35 Transmembrane 140 - 156 ( 139 - 157)  
INTEGRAL Likelihood = -3.03 Transmembrane 55 - 71 ( 53 - 74)  
INTEGRAL Likelihood = -3.03 Transmembrane 247 - 263 ( 246 - 264)  
INTEGRAL Likelihood = -1.12 Transmembrane 393 - 409 ( 393 - 409)  
35 INTEGRAL Likelihood = -0.11 Transmembrane 165 - 181 ( 165 - 181)

----- Final Results -----  
bacterial membrane --- Certainty=0.7050 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
40 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA04279 GB:D17462 Ma+ -ATPase subunit J [Enterococcus hirae]  
Identities = 168/466 (36%), Positives = 260/466 (55%), Gaps = 26/466 (5%)

45 Query: 6 NKTSPFLKSVTQRLTFSPAIVILIGTLLSMPTTHYQNGPNTVLDHFRVNVSVCTG 65  
NK+ K LS Q + F I+IL G LL++FF ++G +T +D F S VCVTG  
Sbjct: 4 MKRVKRVKLSFVQLIAGQFLLILPGGSLTLFFPS-RGSESTHPIDALFATSAVCTG 62

50 Query: 66 LSVVPVAVVNGIQQTIAHAIQICGLVTLINVSFAL-KRMRLSDCTLLGALNRG 124  
L+ + AS +N GQ + M L+IG LG + I+ FA+ K+K+ S +L+ ALN  
Sbjct: 63 LTTNTASVWNSAQQLITLILIEIGLGF-MFILFPAIAKKKISFSMRIVLKRANLE 121

55 Query: 125 DSKDLKHLFFAFKVTFFSLZAFARIVIMIDFIPRGWNGIPIFNSIPLAVSAPCNAGFDNL 184  
+ + + + K ++ A+ + + FIF FGV GI+ SIF AVS+PCNDGFD L  
Sbjct: 122 EMSGVIKLMIYILKFAVVIQVIGAVLSVVFIFERFGKGIWFSIHAVVSFCNDGFDLL 181

Query: 185 GSSSLKDFMLNFTLNVIITFLISGGLGFVAVWDLGVAFKKYFFERPHCYGATFRKLSN 244  
G S L D N L +++++ LII+GGLGF VM D+ +++ + +K++  
Sbjct: 182 GDSSLAD-QTNVYLIMVVSALIIAGGLGFIVWRDI-LSYHR-----VKKTLIL 227

60 Query: 245 SRVLQTTAVILFLGTFVWLEKINSKTIANSFLHQQLMVSFPQTVMRTAGHATISYN 304  
S++ L TR++L +G F+ + + N T+ + ++L +FF +VT RTAG+ +I Y  
Sbjct: 228 SKVALSVTALL-IGGFILFLITERNGTLVKGTPTTERLANTFVMSVPTRTAGYSIDYL 286

65 Query: 305 DTLAPTNIYMIQVIGGAPGGTAGGKIVTTAIIITFLKFAELSGSQSVTFHRIIANKT 364

-1756-

IL M M IGG G TAGG+K TT I + A G++ R I  
 Sbjct: 287 QMSHAGLILMFLMAYIGTSGSTAGGLKTTTLGILLIQHMFGRKTRAEFGRTIRQA 346

5 Query: 365 IKQMTVTLVFFFAVLMIGFILLSSVEPHIAPIP----LLFESISAIATGVSMDLTQGLS 420  
 + + +T L F L + I+LVS I + FE SA TVG++M LTP L+  
 Sbjct: 347 VLRAALT-LFFVTLSLGVVIMVLSVTETIPKTSIGVIEAFVFSAPGTGVLGMLGTLPLAT 405

Query: 421 TAGRLIVIVLMFVGRVGPITVLISLI---QRKEKTIQYATTDILVG 463  
 G+L++I LM++GRVG +TV++SL+ R E +Y I++G  
 10 Sbjct: 406 LIGKLVIISIMYIGRVGIMTVVLSLVKANRAEANYKYPESSIMLG 451

An alignment of the GAS and GBS proteins is shown below.

Identities = 275/462 (59%), Positives = 351/462 (75%), Gaps = 1/462 (0%)

15 Query: 2 GASMGHFFDYKTMVARKLSISFIAVILLGSHILLSLPIFYANAPKTHYIDHLFTTVSMV 61  
 G +MK F K++SV ++L+ SF VIL+G++LLG+P Y N P T Y+DH F VSMV  
 Sbjct: 3 GGMKGRSF-IKSLSVTQRLTFSFAIVILIGTLLSGMPFTHYQNGPNTVYLDHFFNVSMV 61

20 Query: 62 CVTGLSVFFPSKVYNGWQIVAILLMQTGGLGVLTMSLSYTYLRLRMISNDQTLQSAI 121  
 CVTGLSV P+++VYNG GQ ++L+ LMQ G LGLVTL++S + L+RM L+DQTLQSAI  
 Sbjct: 62 CVTGLSVFVFAEVYNGIGQITIANALMQIGLGLVTLIAVSTFALKRKNRISDQTLQSAI 121

Query: 122 TYNSSDCLKKYLIMFKVITLLEVAASILAIDIPRFGLRGKIFNSIFLAVSAFCNAGF 181  
 S DLK YL+ +KVT +LE AA ++ IDPIPRFG +GIFNSIFLAVSAFCNAGF  
 25 Sbjct: 122 NRQDSKDLKHYLFAYKVTFSLEAFIAIVIMIDIPRFQWKNIGNSIFLAVSAFCNAGF 181

Query: 182 DNLKLSLAQKLNPLVNIIVCFLLIISGSLGFVWVKDLIEATQTSHGPKLITKPFERL 241  
 DNL ++SL F LNP +N+I FLIISGSLGFVW DL A + + P ++L  
 30 Sbjct: 182 DNLGSSSLKDFMLNPTLNVITFLIISGSLGFVWVDLQVAFKPYFPERPHCYGATFERL 241

Query: 242 SNHSLVLKLTITILLMQTLLSWLLSPNFRTIANLSLPKQLMVSFPQTVIMRTAGFTI 301  
 SN S+LVL+TT +IL GT L+L LE N +TIAN SL +QLMVSFPQTVIMRTAGF+TI  
 Sbjct: 242 SNQSLRLVLTAVILFLGTFLTWLEKDKNSKTIANFSLQQLMVSFPQTVIMRTAGFTI 301

35 Query: 302 DYTQTDPATNLVYIIMQLIGAPGGTAGGFKVTVIAIILLFLKAEISGQSQVIFHYRTIP 361  
 Y T TN++Y+IQM+IGGAPGGTAGG KVT AI LLKFAELSGQS+VTF R I  
 Sbjct: 302 SYNDTLAPTNLNMIMQVIGGAPGGTAGGKVTIAITFLFLKAEISGQSEVTFERNIA 361

40 Query: 362 SSIKQTLISLITFFFIILISYLLLELNPHIDPFSLFPEASALATGVVMNTNQUTL 421  
 + IKQT++L FFF +L+ G++LL + PH P L FE+ SA+TVGVN+M T Q+  
 Sbjct: 362 NKTIKQTVLVLIFFFAVLMIGFILLSSVEPHIAPILLFESISAIATGVSMDLTQGLST 421

Query: 422 GGRIVIMFLMFIGRVGPITVLSILQKKEIHYAETIILG 463  
 GR++++ LM+GRVGPITVL+S++Q+KKE I YA T+I++G  
 45 Sbjct: 422 AGRILIVIVLMFVGRVGPITVLISLIQRKEKTIQYATTDILVG 463

A related GBS gene <SEQ ID 8851> and protein <SEQ ID 8852> were also identified. Analysis of this protein sequence reveals the following:

Lipid: Possible site: -1 Crend: 9  
 Mcg: Discrim Score: 0.86  
 GVH: Signal Score (-7.5): 0.64  
 Possible site: 45  
 >>> Seems to have a cleavable N-term signal seq.  
 50 ALOM program count: 9 value: -10.14 threshold: 0.0

|    |                      |                     |               |                        |
|----|----------------------|---------------------|---------------|------------------------|
| 55 | INTSGRAL             | Likelihood = -10.14 | Transmembrane | 371 - 387 ( 362 - 391) |
|    | INTSGRAL             | Likelihood = -7.48  | Transmembrane | 200 - 216 ( 190 - 217) |
|    | INTSGRAL             | Likelihood = -4.94  | Transmembrane | 425 - 441 ( 423 - 446) |
|    | INTSGRAL             | Likelihood = -4.67  | Transmembrane | 327 - 343 ( 325 - 349) |
|    | INTSGRAL             | Likelihood = -3.77  | Transmembrane | 81 - 97 ( 81 - 98)     |
| 60 | INTSGRAL             | Likelihood = -2.66  | Transmembrane | 140 - 156 ( 139 - 157) |
|    | INTSGRAL             | Likelihood = -1.33  | Transmembrane | 55 - 71 ( 53 - 71)     |
|    | INTSGRAL             | Likelihood = -0.27  | Transmembrane | 247 - 263 ( 247 - 263) |
|    | INTSGRAL             | Likelihood = -0.11  | Transmembrane | 165 - 181 ( 165 - 181) |
|    | PERIPHERAL           | Likelihood = 2.49   | 308           |                        |
| 65 | modified ALOM score: | 2.53                |               |                        |



-1758-

## Example 1575

A DNA sequence (GBSx1669) was identified in *S.agalactiae* <SEQ ID 4869> which encodes the amino acid sequence <SEQ ID 4870>. This protein is predicted to be TrkA (ktrA). Analysis of this protein sequence reveals the following:

```

5      Possible site: 19
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----
10      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GF:NAC46144 GB:AF001974 putative TrkA [Thermococcus bacter
      ethanolicus]
      Identities = 69/177 (38%), Positives = 110/177 (61%), Gaps = 2/177 (1%)

      Query: 8  VLGLGIFGQTLAQELSNFEQDVTAIDSNPEN--VQVAEVVTKAIGDITDLAFLKHIGI 65
      V+GLG FG +L+L L DV+ ID + E VQA+ +VT A D TD LK + +
20      Sbjct: 6  VLGLGSFGISLAKTLTYMGNDVLVIDESEEELVQAMNGLVTHAVRADATDENVLKSIRV 65

      Query: 66  SDCTVTIATGNSLESSVLAVMHCKKLGVPQVIKARNLVYEEVLYIGADLVISFERES 125
      + D I+A G ++ESS++ M K+IGV VIAKA N ++ VLY++GAD V+ FE++
25      Sbjct: 66  KNFDVAIVAIQNMESSIMVTLVKSLGVCKVIAKHNELHARVLYKVGADRVVMEKDM 125

      Query: 126  GQNVAAANLMNKITDVPQIESDISVIRKFKPKSVGKTVQELNIRHKFDNLNIGIRK 182
      G VA N+ + + D+ + + S+ E + W GKT+++N+R K+ LN++ ++K
      Sbjct: 126  GIKRVARNVPSNLIDILFESKEYSLASLPIFESWFGKTLKEINVRKGYLNVAVKK 182

```

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4715> which encodes the amino acid sequence <SEQ ID 4716>. Analysis of this protein sequence reveals the following:

```

      Possible site: 20
      >>> Seems to have an uncleavable N-term signal seq

35      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 132/221 (59%), Positives = 176/221 (78%)

      Query: 1  MKTKLIGVLGIGFGQTLAQELSNFQDVTAIDSNPENVQVAEVVTKAIGDITDLAFL 60
      + K K +GVIGLIGIFG+T+A+ELSNF+QDVTAID +V+ VA++VTKAA+GDITD FL
45      Sbjct: 2  LKRTKVTVGLGIGFGRTVARELSNFDQDVTAIDIRESHVKEVADLVTKAAVGDITTKFEL 61

      Query: 61  KHIGISDCTVTIATGNSLESSVLAVMHCKKLGVPQVIKARNLVYEEVLYIGADLVIS 120
      +GI CDTV+IA+GN+LBSVLAVMHCKKLGVP +IAKA+N ++EEVLY IGA VI+
50      Sbjct: 62  LAVGLEHCDCTVVIASGNLBSVLAVMHCKKLGVPVTIAKAKNKIPFEEVLYGIGATKVIT 121

      Query: 121  PERESQNVAAANLMNKITDVPQIRSDISVIRKFKPKSVGKTVQELNIRHKFDNLNIGI 180
      PER+SG+ VA+NL+R I + +E IS+IEF IPKSN G+++ +L++R K++LN+IG+
      Sbjct: 122  PERDSGRKVASNLRRHIESIITYLEHGISMKIFKVIKPSWBGQGLSELDVRRKYELNVIGM 181

80      Query: 181  RKAKNKPVDTVEVPINSFLEGGIILVAIANSDAFQRYDLYGY 221
      R+ + K +D+T V PLE I+VAIAN F+++DLYGY
      Sbjct: 182  RQKEVKTLDTNVKPFKFLPWTIIIVAIANDETPEKFDLYGY 222

```





-1760-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1576

A DNA sequence (GBSx1670) was identified in *S. agalactiae* <SEQ ID 4871> which encodes the amino acid sequence <SEQ ID 4872>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -11.62 Transmembrane 73 - 89 ( 68 - 96)
INTEGRAL Likelihood = -11.30 Transmembrane 254 - 270 ( 248 - 274)
INTEGRAL Likelihood = -4.73 Transmembrane 127 - 143 ( 124 - 144)
INTEGRAL Likelihood = -4.19 Transmembrane 50 - 66 ( 47 - 67)
INTEGRAL Likelihood = -3.29 Transmembrane 25 - 41 ( 25 - 45)

----- Final Results -----
bacterial membrane --- Certainty=0.5649 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8855> which encodes amino acid sequence <SEQ ID 8856> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1 Crend: 9
MoG: Discrim Score: -10.49
GVH: Signal Score (-7.5): -1.14
Possible site: 40
>>> Seems to have no N-terminal signal sequence
ALOM program count: 5 value: -11.62 threshold: 0.0
INTEGRAL Likelihood = -11.62 Transmembrane 73 - 89 ( 68 - 96)
INTEGRAL Likelihood = -11.30 Transmembrane 254 - 270 ( 248 - 274)
INTEGRAL Likelihood = -4.73 Transmembrane 127 - 143 ( 124 - 144)
INTEGRAL Likelihood = -4.19 Transmembrane 50 - 66 ( 47 - 67)
PERIPHERAL Likelihood = 3.76 201
modified ALOM score: 2.82

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.5649 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CA813178 GB:Z99110 ykoC [Bacillus subtilis]
Identities = 61/226 (26%), Positives = 108/226 (46%), Gaps = 12/226 (5%)

Query: 49 FLIVVSLGSLVLRKAKIKWQQVSPWILVVVFAVINIMVLFAPHYQDKIYCGSSILL 108
F I+ + G L+ + KW + + F +L V+ A K+ + L
Sbjct: 36 FYIIIVAGVLLAAGILPLKKV-----LLFTIPPLILAFGCWVTAAPV--GRVETTPDNFL 87

Query: 109 KGIGFYDVTSQELFYLENILKYFCIVPLALLFLMTNPSCFASLL-NGLGLSYKIAYAV 167
GP + S + + +L + C L+ ++P+ TT+P F SL Q LS K+AY V
Sbjct: 88 PCAGPISINSQNVSVGISLGFRLCPALSQNVFVTTDPILFMLSVOOCLSPKLAGYV 147

Query: 168 SLTLRYIPDVQREFYITIRACQARGILSKSNLVARIKGNLQIVTPLIPSLERIDTVA 227
R++P +++E I++A + RG + +S ++ +I + P++ S++ + + A
Sbjct: 148 IAGFRFLPLKKQKVLQQAHKIRGG--AARSGINKISALKRYTIPILASAIRKERT 205

Query: 228 TAMELRFRFGKRRRTYWSKQSKSLSDIVLILALASLFSVLYLIHL 273
AME + F ++ RT+Y S+ + D V L L LF +L+ L
Sbjct: 206 LAMESKGFSGRRRTYRTLSVNRERWVFFCLVIL-LFAGSFLVSL 250

```

-1761-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1577

A DNA sequence (GBSx1671) was identified in *S. agalactiae* <SEQ ID 4873> which encodes the amino acid sequence <SEQ ID 4874>. This protein is predicted to be cobalt ABC transporter, ATP-binding protein (cbiO). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.91 Transmembrane 436 - 452 ( 435 - 452)

----- Final Results -----
bacterial membrane --- Certainty=0.1765 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP: CAB13179 GB: Z99110 similar to cation ABC transporter
(ATP-binding protein) [Bacillus subtilis]
Identities = 151/483 (31%), Positives = 246/483 (51%), Gaps = 19/483 (3%)

Query: 8 KDFTFQYDVQSEPTLKGINLSIPKGGKVLILGSPSGSGKSTLGHCLANGIIPNTHKQYSGI 67
+ +F Y+ +P + I+ + KGE VL+LGPSG GKS+L CLANG+ P G SG
Sbjct: 11 EQLSFYSEDEKPVFQDISFELQKGEVLLLGSPSGGKSSALCLANGLYPEACDGIQSGH 70

Query: 68 FTINHQIAFDLSIYDK-SHLVSTVLQDPDQGFGLTVAEIDAPALENDVVAQERMAISIVE 126
+ K D + + + V QDPD QF LTV ++IAP LEN + +EEM +
Sbjct: 71 VFLFQKPVITDAETSETITQHAGVVPQDPDQGFQCMLTVEDEIAPGLENLQIPKSEMTKIN 130

Query: 127 MNAKRLIEIAPLLSKRFQDLSGGGQKQVSLAGVLVDDSPILLFDEPLANLDQPSGGQDIMAL 186
+ L I L K LSGGQKQ+V+LA +L + +++ DEP + LDP S ++ + L
Sbjct: 131 AVLGKLRITHLKERNISTLSGGGQKQVLAACILAMEPELITLDEPTSLDPPSAREFVHL 190

Query: 187 VDRHQEQDATTILIEHRLD--VFYERVDRVVLPSDGGQIYNGEPDQLL--KTNPLSEY 242
+ + +E+ + ++IRH+L++ + ER +VL G+ +G L + L +
Sbjct: 191 MKDLQREKGSLLVIEHQLEDEWAFWERT--IVLDKSGKALDGLTGNLFQHEAETLKKL 246

Query: 243 GIREFLYISALKNLGYDFEKQNTMTSIDDFPSELLIPMRALDLDKHTDKLLSVQHLV 302
GI P + L F M + + K+A + +L V LS
Sbjct: 249 GIAIPVCHLQKESMPFTLSKEMLPKEPIAGH--VKKKA---PSGESVLEVSLSF 302

Query: 303 SYDLENNITLDDVSFDLYKGQRLAIVGKNGAGKSTLAKALCOFI-PNNATLYNNEVSDQ 361
+ + D+SF L+G A+VG NG GKSTL L + P + + + + +
Sbjct: 303 ARG-QQAIFKDISPSLRBGLSLTALVQPHOTGKSTLLSVLASLMKPGSGKILLDQPLQKY 361

Query: 362 SIKERASRIGYVLQNPQMSIQAMVDFEVALGLRLRGPSDNDIESRVYDILKVCGLYQFR 421
KE +R+G+V QNP V+DE+ G+ ++ + E+ +L+ GL
Sbjct: 362 KEKELRKRMGFVQNPQHEQFVTDVYDELLFGQK---ANAETEKKQHLQRLPGLAHLA 417

Query: 422 NWPISALSPGQKKRVTIASILILNPEVILDEPTAGQDMKHYTEMMSFLDKLSCOGHETIV 481
+ +A+S GQK+R+++A++L++ +V++LDEPT QD + E M + ++ +G ++
Sbjct: 418 DHHPFATISOGCKRRLSVATVMAHIDVKVLLDDEPTQGDARTAECEMIQRIKAGSTAVL 477

Query: 482 MIT 484
MIT
Sbjct: 478 MIT 480
```

There is also homology to SEQ ID 4416.

SEQ ID 4874 (GBS424d) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 2 & 4; MW 77kDa) and in Figure 239 (lane 10; MW 77kDa). It

-1762-

was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 5 & 7; MW 52kDa) and in Figure 182 (lane 4; MW 52kDa). Purified GBS424d-His is shown in Figure 241, lanes 6 & 7. Purified GBS424d-GST is shown in Figure 246, lane 12.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1578

A DNA sequence (GBSx1672) was identified in *S.agalactiae* <SEQ ID 4875> which encodes the amino acid sequence <SEQ ID 4876>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -8.12    Transmembrane    39 - 55 ( 35 - 63)
INTEGRAL    Likelihood = -3.98    Transmembrane    72 - 88 ( 71 - 90)
INTEGRAL    Likelihood = -3.66    Transmembrane   108 - 124 ( 106 - 127)
INTEGRAL    Likelihood = -2.34    Transmembrane   182 - 198 ( 181 - 198)
INTEGRAL    Likelihood = -1.44    Transmembrane   141 - 157 ( 139 - 158)

----- Final Results -----
bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE59830 GB:AJ012388 hypothetical protein [Lactococcus lactis]
Identities = 109/182 (59%), Positives = 141/182 (76%)

Query: 31 MNTNITKKVVATGIGAAFLPIIGMLVNIPTPIPTNITQLQYAVLAFVITYGPGVGFPTG 90
M N+K VVATGIGAAFL+IIG L+NIPTPIPT+IQLQYAVLALF+ ++GF GF G
Sbjct: 1 MKNSVKIVVATGIGAAFLVIIGMLINIPTPIPTNTS IQLQYAVLALFSAFGPLAGFLIG 60

Query: 91 FIGHALKDSIQYGSFWWTWVLVSGLLGLMIGFFAKKLAIQLSGMTKKDLLFNVVQVIAN 150
FIGHALKDS YG+FWWTWVL SGL+GL +GF K+ ++ K++ FN+VQ +AN
Sbjct: 61 FIGHALKDSFLYQAPWWTWVLVSGLLGLFLGFGVKRESLQGFQGNKKIIRFNI VQFLAN 120

Query: 151 LIGNSVVPYDGIDIFFYSEPAKVKFAGGFLSSLVNSITIGVGGTLLLLAYAKSRPQKGSLS 210
++ W +AAP GDI YSEPA+KVF QG ++ LVN++TI V GTLLL YA -R+ G+L
Sbjct: 121 VVVWGLIAPIGDILVYSEPAKVKFTQGVVAGLVNALTIAVAGTLLLLKLYAATRTKSGTLD 180

Query: 211 KD 212
K+
Sbjct: 181 KE 182

```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8857> and protein <SEQ ID 8858> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 6
McG: Discrim Score: -5.01
GVH: Signal Score (-7.5): -5.9
Possible site: 50
>>> Seems to have no N-terminal signal sequence
ALOM program count: 5 value: -8.12 threshold: 0.0
INTEGRAL    Likelihood = -8.12    Transmembrane    31 - 47 ( 27 - 55)
INTEGRAL    Likelihood = -3.98    Transmembrane    64 - 80 ( 63 - 82)
INTEGRAL    Likelihood = -3.66    Transmembrane   100 - 116 ( 98 - 119)
INTEGRAL    Likelihood = -2.34    Transmembrane   174 - 190 ( 173 - 190)
INTEGRAL    Likelihood = -1.44    Transmembrane   133 - 149 ( 131 - 150)
PERIPHERAL  Likelihood = 5.78
modified ALOM score: 2.12

```

-1763-

\*\*\* Reasoning Step: 3

----- Final Results -----

5           bacterial membrane --- Certainty=0.4248(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10   ORF02330 (367 - 912 of 1212)  
     GP|6165407|emb|CAB59830.1||AJ012388 (1 - 182 of 182) hypothetical protein {Lactococcus  
     lactis}  
     %Match = 28.1  
     %Identity = 59.9 %Similarity = 78.6  
     Matches = 109 Mismatches = 39 Conservative Sub.s = 34

15           102       132       162       192       222       252       282       312  
     MOVVGQFIVGVQDSCETALNSSTDVLFITAVAKSVFGKK\*TNEGRLYSI\*DLFWYLIFSIVQFPLSIRPQISLKYD

20           342       372       402       432       462       492       522       552  
     KIEQIVSDCLSLFFREVFMTNTIKKVATSIGRALFIIGMLVNIPTPIPTNTNIQLQYAVLALFAVIGFGVGFPTGI  
     |   :::|   |||||   |||||   :::|   :::|   |||||   |||||   :::|   :::|   :::|  
     MKNNSVKIVVATSIGRALFVIIGWLNIPPTPIPTNTSIQLQYAVLALFSAIFLAGFLIGFI  
                   10       20       30       40       50       60

25           582       612       642       672       702       732       762       792  
     GHALKDSIQVGSFWTWVLSGLGLMIGFPAKLAIQLSGMTKDLLLFVQVQVIANLIGNSVVAIFYGDIFFYSEPAK  
     |||||   ||:|||||   ||:|   ::|   :::   ||:|   ::|   ::|   ||:|   ||:|   ||:|  
     GHA\*KDSFLYGAPWVWVLSGLMLFLGFGVKRESLTQIFGNKEIIRFNIVQFLANVWVWGLIAPIDILVYSEPAK  
                   80       90       100       110       120       130       140

30           822       852       882       912       942       972       1002       1032  
     VFAQGLSSLVNSITIGVGGTLLLLAYAKSRPQKSLSKD\*DKRVIYERFY\*MBGFYLSI\*RSI\*TNFKRD\*LKHS\*R\*K  
     ||   ||   ::   ||:|   ||   |||   ||   ::|   ::|   ||   ||  
     VFTQGVVAGLVKALTAVAGTLLLLKLYATRTKSTLDKE  
                   160       170       180

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40   **Example 1579**

A DNA sequence (GBSx1673) was identified in *S.agalactiae* <SEQ ID 4877> which encodes the amino acid sequence <SEQ ID 4878>. Analysis of this protein sequence reveals the following:

Possible site: 42

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

45   INTEGRAL   Likelihood = -6.85   Transmembrane   86 - 102 ( 80 - 106)

----- Final Results -----

50           bacterial membrane --- Certainty=0.3739(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

55

## Example 1580

A DNA sequence (GBSx1674) was identified in *S. agalactiae* <SEQ ID 4879> which encodes the amino acid sequence <SEQ ID 4880>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have a cleavable N-term signal seq.
5  INTEGRAL    Likelihood = -3.61    Transmembrane    107 - 123 ( 96 - 124)
   INTEGRAL    Likelihood = -1.86    Transmembrane    124 - 140 ( 124 - 142)
   INTEGRAL    Likelihood = -1.38    Transmembrane    83 - 99 ( 83 - 100)
   INTEGRAL    Likelihood = -1.12    Transmembrane    142 - 158 ( 142 - 160)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.2444 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9415> which encodes amino acid sequence <SEQ ID 9416> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AACT6124 GB:AR000391 putative transport protein [Escherichia
20  coli K12]
   Identities = 139/178 (78%), Positives = 159/178 (89%)

Query: 1  MVGIMLVVALVNPILAFVMMRQIPVLVLRCLKDSGITAFFTRSSAANIPIVMRLCEDL 60
   +VG ML VALVNP++ + +R+NP+PLVL CL+SG+ AFFTRSSAANIPIVM LCE L
25  Sbjct: 222 LVGCMILVALVNPILVNWVKIRNPFFLVLLCLRESGVYAFFTRSSAANIPIVMALCEKL 281

Query: 61  GLDKDTYSVSIPLAGAINMAGAAITINIL/LAAVNTLIGITVDFPTAFLLSVAAVSACGA 120
   LD+DTYSVSIPLAGA INMAGAAITI +LTLAAVNTLGI VD PTA LLSVA++ ACGA
30  Sbjct: 282 NLDRDTYSVSIPLAGATINMAGAAITITVTLAAVNTLGIPLDLPTALLSVASLACAGA 341

Query: 121  SGVITGSGLLLIIPVACSLFGISNDVAMQVVGVIQVQDSCTALNSSTDVLPTAA 178
   SGV GGSLLLIIP+AC++FGISND+AMQVV VGF+GV+QDSCTALNSSTDVLPTA A
35  Sbjct: 342 SGVAGSGLLLIPLACNMFGISNDIMQVAVGFIIGVQDSCTALNSSTDVLPTAA 399

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4881> which encodes the amino acid sequence <SEQ ID 4882>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have an uncleavable N-term signal seq
40  INTEGRAL    Likelihood =-13.69    Transmembrane    212 - 228 ( 202 - 239)
   INTEGRAL    Likelihood = -7.38    Transmembrane    78 - 94 ( 74 - 108)
   INTEGRAL    Likelihood = -6.53    Transmembrane    179 - 195 ( 175 - 200)
   INTEGRAL    Likelihood = -6.10    Transmembrane    315 - 331 ( 312 - 341)
   INTEGRAL    Likelihood = -5.36    Transmembrane    44 - 60 ( 42 - 61)
   INTEGRAL    Likelihood = -4.41    Transmembrane    13 - 29 ( 11 - 41)
45  INTEGRAL    Likelihood = -3.19    Transmembrane    340 - 356 ( 333 - 358)
   INTEGRAL    Likelihood = -3.08    Transmembrane    145 - 161 ( 144 - 162)
   INTEGRAL    Likelihood = -0.90    Transmembrane    358 - 374 ( 358 - 376)

----- Final Results -----
50  bacterial membrane --- Certainty=0.6477 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

55  >GP:AAF95950 GB:AR004347 sodium/dicarboxylate symporter [Vibrio cholerae]
   Identities = 243/385 (63%), Positives = 299/385 (77%), Gaps = 2/385 (0%)

Query: 9  VRVSLIKKIGIGVIGVIMLGILAPDVG-FSLIGKLVGGLKATAPLLVPALVNSQAISHQ 67
   VR +L+ +I G++G + +P+ ++G LFGV LKA+AP+LVF LV+ I++Q

```

-1765-

Sbjct: 11 VRGNLVLIQILGILGAAAMATFSPKYAQKVLIGNLFVGLKAVAPVILVFIIVASSIANQ 70  
 Query: 68 KKGQTNMTLITVLVLFGTFSALAVLTAYLFPITVLNTPVNTLSPPQGVAVPQSL 127  
 KK + T M I+VLYLFGTF++AL AV+ ++LFP TIVL T ++PQG+ARV +L  
 5 Sbjct: 71 KKNQHTYMRPIVLYLFGTFSAALTAIVLSLFPITLVLATGAEGA-TPQGIARVILNTL 129  
 Query: 128 LKLVADNPINALATANYIGVLSWAIIRGLAKAASKETKHLIKTAAREVTSQIVVMIINLA 187  
 L KLVADNP++AL ANYIG+L+W + GLAL +S TK + + + SQIV +II LA  
 10 Sbjct: 130 LKFLVDNPVSALMANYIGILAWGVGLALHISSTTKAVFEDLSHGISQIVRPIIRLA 189  
 Query: 188 PIGIMSLVFTTISENGVILSDYAFILVLVMTLFFVALVNPVLAIVLTITQNPIPLVLR 247  
 P GI LV +T + G L+ YA L+ VL+G M P+ALVNP+I + R+N+PLVL+  
 Sbjct: 190 PFGIFGLVASTFATTGFDALAGYAQLLAVLLGMAFIALVNMHIVYKIRNPFPFLVLIQ 249  
 Query: 248 CLRESGLTAFFTRSSAANIPIVMMLCKKLGKSDTYSVSIPLGATINMGGAITINVLTL 307  
 CLRESG+TAFFTRSSAANIPIVM LCK+ L +DTYSVSIPLGATINM GAAITI VILT  
 15 Sbjct: 250 CLRESGVTAFFTRSSAANIPIVMALCKKLDEDTYSVSIPLGATINMAGAAITITVLT 309  
 Query: 308 AAVHTFGIPIDFLTALLSVVAASACGASGVAGGSELLIPVACSLFGISNDLMAQVGV 367  
 AAVHT GI +D +TALLSVVAASACGASGVAGGSELLIP+AC LPIGIND+AMQIV V  
 20 Sbjct: 310 AAVHTMGIEVDLTALLSVVAASACGASGVAGGSELLIPACGLFGISNDIAMQVAV 369  
 Query: 368 GFIVGVIQDSCETALNSSTDVLF 392  
 GF1+GVIQDS ETALNSSTDVLF 392  
 25 Sbjct: 370 GFIVGVIQDSCETALNSSTDVLF 394

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/186 (82%), Positives = 172/186 (92%)

30 Query: 1 MVGTMLFVALVNPPIIAFVMMKPNPYPLVLRCLKDSGITAFFTRSSAANIPIVMRLCEDL 60  
 +VGTMFLFVALVNP+IA ++ R+NPYPLVLRCL++SG+TAFFTRSSAANIPIVM+LC+ +  
 Sbjct: 217 LVGTMLFVALVNPVLAIVLTITQNPIPLVLRCLRESGLTAFFTRSSAANIPIVMMLCKKI 276  
 Query: 61 GLDKDTYSVSIPLGAAINMAGAAITINILLTAARVNTLGITVDPTAFLLSVVAASACGA 120  
 GL KDTYSVSIPLGA INM GAAITIN+LTAARV+T GI +DF TA LLSVVAASACGA  
 35 Sbjct: 277 GLSKDTYSVSIPLGATINMGGAITINVLTLAARVHTFGIPIDFLTALLSVVAASACGA 336  
 Query: 121 SGVTGSGLLLIIPVACSLFGISNDVAMQVGVGFIVGVIQDSCETALNSSTDVLF 180  
 SGV GGSLLLIIPVACSLFGISND+AMQVGVGFIVGVIQDSCETALNSSTDVLF+AE  
 40 Sbjct: 337 SGVAGGSELLIPVACSLFGISNDLMAQVGVGFIVGVIQDSCETALNSSTDVLF 396  
 Query: 181 SVPGKK 186  
 + + +K  
 45 Sbjct: 397 AFWKK 402

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1581

A DNA sequence (GBSx1675) was identified in *S. agalactiae* <SEQ ID 4883> which encodes the amino acid sequence <SEQ ID 4884>. This protein is predicted to be acid phosphatase. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2436 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not clear) < succ>

-1766-

A related GBS nucleic acid sequence <SEQ ID 9427> which encodes amino acid sequence <SEQ ID 9428> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
   Identities = 167/251 (66%), Positives = 209/251 (82%)

Query: 7  BQKTKFKNISLSSNKLIAKENTMSVINYQNGARAKALYLQGYNAKMLDMLQKPSKD 66
++ K ++ S ++ L + ENIMSVIHWQ +AKAKALVIQGY +A +L + L + ++KP
10 Sb|ct: 34 KETVVGQTKVITYSDQLRSENENTMSVINYQRAAEAKALYLQGYLATDRLKNGLQQAQTKD 93

Query: 67 YSIILDIDETVLNDSFYQAKNIKDGGSPFPESWDKWNQKKKAKAVAGAKEFLKYANKEGI 126
YSI+LD+DETVDLNSFYQAKNI +G+SFTFPESWD NVQKK AK VAGAKEFL+AA+ G+
15 Sb|ct: 94 YSIVLDIDETVLNDSFYQAKNILEGTSFTFPESWDVWVQKKEAKPVAGAKEFLQPADQNGV 153

Query: 127 KIYVSDRTDAQVDATKENLEKEGIPVQGDHLLFLKKEKMSKESRRQAQKDLNLMFL 186
+IYY+SDR +QVDAT ENL+KEGIPVQGDHLLFL+G+KSKE+RQ V+ +NLNLMFL
15 Sb|ct: 154 QIYYISDRAVSQVDATMENLQKEGIPVQGDHLLFLREGVKSKEARRQXVKETINLMFL 213

Query: 187 GGNLVDPADPSKSSSTREQLLTKLQSEFGSKPIVF+FNPMYGSWESA+Y+G LD OL 246
GNLVDPADPSK S DR LL++LQ EFG -FI+FNPMYGSWESA+Y+G LD OL
20 Sb|ct: 214 GGNLVDPADPSKSSSEDRALLSELQREFGRQFIFFNPMYGSWESAIVYKDKLDASQQL 273

Query: 247 KERQGNLHSDYD 257
KER+K L S++
25 Sb|ct: 274 KERRIALESFE 284

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4885> which encodes the amino acid sequence <SEQ ID 4886>. Analysis of this protein sequence reveals the following:

```

Possible site: 25

30 >>> May be a lipoprotein

----- Final Results -----
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
35          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

40 >GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
   Identities = 234/284 (82%), Positives = 261/284 (91%)

Query: 1  MKSEKVVSVISLITLSLFLVTGCAKVDNNKSVNLKPKATQKTYNSYDDQLRSRENTSVLW 60
MK+K+V SVISL LSLFLVTGCA++D+ +VN K KOT +YSD+QLRS ENIMSVLW
45 Sb|ct: 1  MKTKQVASVISLISLFLVTGCAQLDHKANVNSKETVQKTKVITYSDQLRSENENTSVLW 60

Query: 61 YQRAAETQALYLQGYQLATDRLEKQLNKPTDKPYSIVLDIDETVLNDSFYQAKNILEGTS 120
YQRAAE +ALYLQGYQLATDRLEK OL +TDKPYSVILIDETVLNDSFYQAKNI+LEGT
50 Sb|ct: 61 YQRAAKAKALYLQGYQLATDRLEKNGLQQAQTKPYSIVLDIDETVLNDSFYQAKNILEGTS 120

Query: 121 FTFPSWDYVWQKKEAKPVAGAKFIQPADQNGVQIYYISDRSTTQVDATMENLQKEGIPV 160
FTFPSWD VVQKKEAKPVAGAK+FLQPADQNGVQIYYISDR+ +QVDATMENLQKEGIPV
50 Sb|ct: 121 FTFPSWDVWQKKEAKPVAGAKEFLQPADQNGVQIYYISDRAVSQVDATMENLQKEGIPV 160

Query: 181 QGRDHLLFLEKGVKSKESSRRQVKETINVTIMLFGGNLDDPADPSKSSQEDRTALLSDLQE 240
QGRDHLLFLE+GVKSKE+RRQVKETIN+ MLFGGNL+DPADPSKKS+EDRTALLS+LQE
55 Sb|ct: 181 QGRDHLLFLEGVKSKEARRQVKETINVTIMLFGGNLDDPADPSKSSQEDRTALLSELQE 240

Query: 241 BFGRRFIIFNPMYGSWESGAIYKGRKLDVLKOLEERRKSLKSPK 284
BFGR+FIIFNPMYGSWE A+YK+KLD QL+ERRK+L+SF+
60 Sb|ct: 241 BFGROFIIFNPMYGSWESAIVYKGRKLDASHOLKERRKALESE 284

```

An alignment of the GAS and GBS proteins is shown below.



-1767-

Identities = 166/247 (67%), Positives = 207/247 (83%)

Query: 10 TKPQGISLSNKLAKENTMSVLWYQNSAEAKALYLQYTVARMKLLDWLQKPKSEKPSI 69  
 TK S S ++L ++NNTMSVLWYQ ++AE +ADVLQSY +A +L +L KP++KPSI  
 5 Sbjct: 37 TKQTNSYSDDQLRSKENTMSVLWYQRAAZTQALYLQYTVARMKLKBQLNKPTDKPSI 96

Query: 70 ILDLDETVLNSPYQAKNIKDGSSPTPESSWDKWKQKSAKAVAGAKEFLKYANKEGIKIY 129  
 +LD+DETVLNSPYQAKN+ +G+ PTPESSWD WQKK AK VAGAK+FL++A++ G++IY  
 10 Sbjct: 97 VLDIDETVLNSPYQAKNVLBSTGPTPESSWDYWWQKKRAKPVAGAKDFLQFADQNGVQIY 156

Query: 130 YVSDRTDAQVDATKENLEKSGIPVQGDHLLFLKKGKSKSRQAVQKUTHLIMLFGDN 189  
 Y+SDR+ QVDAT ENL+KSGIPVQGDHLLFL+K+KSGKSRQ V++ TN+ MLFGDN  
 15 Sbjct: 157 YISDRTTQVDATMENLQKSGIPVQGRDHLFLKKGKSKSRQKVKETTNVIMLFGDN 216

Query: 190 LVDPADFSKSSSTDRBQLLTKLQSEFGSKFVFPFNMYGSWESAIYQGRHLVQKQLGR 249  
 L+DPADFSK S DR LL+ LQ EFG +FI+FPFNMYGSWE AIY+G+ LDV KQL+ER  
 20 Sbjct: 217 LLDPADFSKSKSEURTFALLSDLQSEFGRRFTIIPFNMYGSWEGAIYKGBLDVLKQLGR 276

Query: 250 QKGLHSY 256  
 +K L S+  
 20 Sbjct: 277 RKSLSKF 283

SEQ ID 9428 (GBS661) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 2 & 4; MW 61kDa + lane 3; MW 27kDa) and in Figure 186 (lane 11; MW 61kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 5-7; MW 25kDa).

GBS661-GST was purified as shown in Figure 237, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 30 Example 1582

A DNA sequence (GBSx1676) was identified in *S. agalactiae* <SEQ ID 4887> which encodes the amino acid sequence <SEQ ID 4888>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succo>  
 40 bacterial membrane --- Certainty=0.0000(Not Clear) < succo>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succo>

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4889> which encodes the amino acid sequence <SEQ ID 4890>. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succo>  
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succo>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succo>

An alignment of the GAS and GBS proteins is shown below.

Identities = 395/398 (99%), Positives = 398/398 (99%)

Query: 1 MAKLTIVKIDVLDLGGKVLVRVDENVPLKDGVTINDNRITAALEPTIKYIIBQGGRAILFSLH 60

-1768-

```

      MAKLTQKDVLDLGGKKVLVRVDPNVPLKDGVTINDNRITAAIPITYKIIEQGGRAILFSLH
Sbjct: 1  MAKLTQKDVLDLGGKKVLVRVDPNVPLKDGVTINDNRITAAIPITYKIIEQGGRAILFSLH 60

      GRVKEADKEGKSIAPVAADLAALKQDVVFPQVTRGAKLREAINALEDDGQVLLVENTRF
5  Query: 61  GRVKEADKEGKSIAPVAADLAALKQDVVFPQVTRGAKLREAINALEDDGQVLLVENTRF 120
      GRVKEADKEGKSIAPVAADLAALKQDVVFPQVTRGAKLREAINALEDDGQVLLVENTRF
Sbjct: 61  GRVKEADKEGKSIAPVAADLAALKQDVVFPQVTRGAKLREAINALEDDGQVLLVENTRF 120

      EDVDGKKSKNDRELQKYWASLGDGIFVNDAPGTAHRAHASNVGISANVEKAVAGFLLEN
10 Query: 121  EDVDGKKSKNDRELQKYWASLGDGIFVNDAPGTAHRAHASNVGISANVEKAVAGFLLEN 180
      EDVDGKKSKNDRELQKYWASLGDGIFVNDAPGTAHRAHASNVGISANVEKAVAGFLLEN
Sbjct: 121  EDVDGKKSKNDRELQKYWASLGDGIFVNDAPGTAHRAHASNVGISANVEKAVAGFLLEN 180

      EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENTLEKADKVLIGGQMTYTFYKAQSIIE
15 Query: 181  EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENTLEKADKVLIGGQMTYTFYKAQSIIE 240
      EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENTLEKADKVLIGGQMTYTFYKAQSIIE
Sbjct: 181  EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENTLEKADKVLIGGQMTYTFYKAQSIIE 240

      GNSLVEEDKLDVADLLEKSKGLILPVDKSRANAFAGYTEVRDTEGKAVSBEGLGLDIG
20 Query: 241  GNSLVEEDKLDVADLLEKSKGLILPVDKSRANAFAGYTEVRDTEGKAVSBEGLGLDIG 300
      GNSLVEEDKLDVADLLEKSKGLILPVDKSRANAFAGYTEVRDTEGKAVSBEGLGLDIG
Sbjct: 241  GNSLVEEDKLDVADLLEKSKGLILPVDKSRANAFAGYTEVRDTEGKAVSBEGLGLDIG 300

      FKSIADFDEALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIIGGDSAAA
25 Query: 301  FKSIADFDEALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIIGGDSAAA 360
      FKSIADFDEALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIIGGDSAAA
Sbjct: 301  FKSIADFDEALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIIGGDSAAA 360

      AINLGRADKFSWLSTGGASMELEGGKVLPLGLAALTEK
30 Query: 361  AINLGRADKFSWLSTGGASMELEGGKVLPLGLAALTEK 398
      AINLGRADKFSWLSTGGASMELEGGKVLPLGLAALTEK
Sbjct: 361  AINLGRADKFSWLSTGGASMELEGGKVLPLGLAALTEK 398

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 30 Example 1583

A DNA sequence (GBSx1677) was identified in *Sagalactiae* <SEQ ID 4891> which encodes the amino acid sequence <SEQ ID 4892>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence
35  INTEGRAL    Likelihood = -8.39    Transmembrane    97 - 113 ( 93 - 118)
      INTEGRAL    Likelihood = -3.66    Transmembrane    25 - 41 ( 24 - 48)
      INTEGRAL    Likelihood = -3.40    Transmembrane    121 - 137 ( 121 - 140)
      INTEGRAL    Likelihood = -3.24    Transmembrane    72 - 88 ( 72 - 88)
40  INTEGRAL    Likelihood = -2.07    Transmembrane    143 - 159 ( 143 - 160)

----- Final Results -----
      bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4893> which encodes the amino acid sequence <SEQ ID 4894>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence
50  INTEGRAL    Likelihood = -8.23    Transmembrane    97 - 113 ( 93 - 118)
      INTEGRAL    Likelihood = -7.17    Transmembrane    121 - 137 ( 119 - 140)
      INTEGRAL    Likelihood = -4.13    Transmembrane    25 - 41 ( 24 - 48)
      INTEGRAL    Likelihood = -3.24    Transmembrane    72 - 88 ( 72 - 88)
55  INTEGRAL    Likelihood = -2.55    Transmembrane    154 - 170 ( 154 - 170)

----- Final Results -----
      bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1769-

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/178 (87%), Positives = 159/178 (94%)

```

5   Query: 1  MRTLKLLSNYKFDIKKFKLGMRTKTGLSVFLVLLVPHLFGWKGLOIGALTAVFSLRED 60
      MRTLKLLSNYKFDIKKFKLGMRTKTGLSVFLVLLVPHLFGWKGLOIGALTAVFSLRED
      Sbjct: 1  MRTLKLLSNYKFDIKKFKLGMRTKTGLSVFLVLLVPHLFGWKGLOIGALTAVFSLRED 60

10  Query: 61  FDKSVHFGFSRIIGNSIGGLSLVFFAFNEIFHQAPWVILLIIVPICIMLCIMINVAQNNK 120
      FDKSVHFGFSRIIGNSIGGLSLVFFAFNEIFHQAPWVILLIIVPICIMLCIMINVAQNNK
      Sbjct: 61  FDKSVHFGFSRIIGNSIGGLSLVFFAFNEIFHQAPWVILLIIVPICIMLCIMINVAQNNK 120

15  Query: 121 SGIIGTAAALLIITLSIPSGETILVVFARIFETFCGVFIAMVNTDIEILKKLQNNK 178
      SGIIGTAAALLIITLSIPSGETILVVFARIFETFCGVFIAMVNTDIEILKKLQNNK
      Sbjct: 121 SGIIGTAAALLIITLSIPTGQFFIYVTSRVFETFCGVFVAILVNTDVELIKMKNFKK 178

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 20 Example 1584

A DNA sequence (GBSx1678) was identified in *S. agalactiae* <SEQ ID 4895> which encodes the amino acid sequence <SEQ ID 4896>. This protein is predicted to be regulatory protein glnr (glnR). Analysis of this protein sequence reveals the following:

```

25  Possible site: 17
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
30          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA00402 GB:D00513 ORF129 [Bacillus cereus]
Identities = 59/123 (47%), Positives = 89/123 (71%), Gaps = 5/123 (4%)

35  Query: 4  RELRRTAVFPFIGAVNKLDTLTARQIRYYEQQLITPERTBGNRRMPSLMDLRLEIKD 63
      +E RR+ +PFIG VM LT L+ARQIRYYE+ L++P RT+GNRR+FS ND+D+LLEIKD
      Sbjct: 2  KEDRSAPLFPFIGVMDLTLSARQIRYYEENHNLVSPTTKGNRRKLSFNDVKLEIKD 61

40  Query: 64  FISDGLHISDIKNEYMQRCH-----KSEKQKSLSDAERVRLQLDELNRQGFSSPSQHI 118
      + GL+++ IK + +++ K KE+ K +S E+R++L+DEL++ GRF+ S
      Sbjct: 62  LLDQGLNAGIKQVLMDKRNQTVRAVKVSETKSEIKTELAKILRDELQHTGRFNRTSLRQ 121

45  Query: 119 GNM 121
      G++
      Sbjct: 122 GDI 124

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4897> which encodes the amino acid sequence <SEQ ID 4898>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 20
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
55          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

-1770-

The protein has homology with the following sequences in the databases:

```
>GP:BA00402 GB:D00513 ORF129 [Bacillus cereus]
Identities = 59/122 (48%), Positives = 83/122 (67%), Gaps = 5/122 (4%)

5 Query: 4 KELRRSMAVFFIGTVMTLTDLRSARQIRYYEDQGLIKPERTQGNRRMFSINMDRLLEIKD 63
KE RRS +FFIG VM LT LRSARQIRYYB+ L+ P RT+GNRR+FS ND+D+LLEIKD
Sbjct: 2 KEORRSAPLFFIGTVMDLTQLRSARQIRYYERHNLVSPYTKGNRRFLPSFNDVDKLEIKD 61

10 Query: 64 FLSEGLNTAAIKRSTVERQG-----KLMQKQKALTADVRRILHDEMLTQSGFSTPSQHI 118
L +GLN+A IK+ ++ K+ ++ K ++ ++R+IL DE+ F+ S
Sbjct: 62 LLDQGLNAGIRQVLAKENQTEAVKVKRETKKISKTSRLKILRDELQHTGRFNRTSLRQ 121

Query: 119 GN 120
G+
15 Sbjct: 122 GD 123
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 90/123 (73%), Positives = 106/123 (87%)

20 Query: 1 MKERELRTHAVFFIGAVNKLTDLTARQIRYYEDQGLITPERTEGNRRMFSINMDRLLE 60
MKE+ELR+NAVFFIG VM LTDL+ARQIRYYEDQGLI PERT+GNRRMFSINMDRLLE
Sbjct: 1 MKEKELRPSNAVFFIGTVMTLTDLRSARQIRYYEDQGLIKPERTQGNRRMFSINMDRLLE 60

25 Query: 61 IKDFISDGLHISDIKNEYMQRQHKSKKKSLSDAEVRRLIQDELNRQGRFSSPSQHIGN 120
IKDF+S+GL+I+ IK EY++RQ K +KQK+L+DA+VRR+L DE+ Q FS+PSQHIGN
Sbjct: 61 IKDFISDGLNIAAIKREYVERQGKLMQKQKALTADVRRILHDEMLTQSGFSTPSQHIGN 120

Query: 121 MHL 123
+
30 Sbjct: 121 FRI 123
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1585

35 A DNA sequence (GBSx1679) was identified in *S. agalactiae* <SEQ ID 4899> which encodes the amino acid sequence <SEQ ID 4900>. This protein is predicted to be glutamine synthetase (glnA). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2157 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4901> which encodes the amino acid sequence <SEQ ID 4902>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence
50 INTEGRAL Likelihood = -0.00 Transmembrane 347 - 363 ( 347 - 363)

----- Final Results -----
bacterial membrane --- Certainty=0.1001 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
55 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

-1771-

Identities = 392/448 (87%), Positives = 421/448 (93%)

Query: 1 MTITAEIDIRREVKEKNVTLRLMPTDILGVMIKNVEIPATDEQLDKVLNSKAMPDGGSSIEG 60  
 5 M IT DIRREVKEKNVTLRLMPTDI+GVMIKNVEIPAT DEQLDKVLNSK MPDGGSSIEG 60  
 Sbjct: 1 MAITVADIRREVKEKNVTLRLMPTDIMGVMIKNVEIPATKEQLDKVLNSKMPDGGSSIEG 60

Query: 61 FVRINESDMYLYPDLDTWIVFPWGDENGVAGLICDIYTABG+PFAGDPGRNKLKRMKRM 120  
 10 FVRINESDMYLYPDLDTWIVFPWGDENGVAGLICDIYTABG+PFAGDPGRNKLK R+K M 120  
 Sbjct: 61 FVRINESDMYLYPDLDTWIVFPWGDENGVAGLICDIYTABGKPFAGDPGRNKLRLALGHM 120

Query: 121 QEMKYKSNLGPPEKFFFLKMDENKNPTLDVNDKGGYFLAPITDLADNTRREIVNVLITQN 180  
 E+GYKSNLGPPEKFFFLKMD+ GNPTL+VND GGYFLAP DLADNTRREIVN+LT+M 180  
 Sbjct: 121 NEIGYKSNLGPPEKFFFLKMDDKGNPTLEVNNDGGYFLAPIDLADNTRREIVNLTQM 180

Query: 181 GFVEASHHEVAVGQHEIDFKYDVLKACINIQFLKLVKTIARKHGLYATPMKPKFGI 240  
 15 GFVEASHHEVAVGQHEIDFKY DVLKACINIQ+FLKLVKTIAR+HGLYATPMKPKFGI 240  
 Sbjct: 181 GFVEASHHEVAVGQHEIDFKYADVLKACINIQI FLKLVKTIAREHGLYATPMKPKFGI 240

Query: 241 NGSQNHCHNMLFDNENGNNAFFDPEDPRGMQLSEDAITFLGLMKHAYNTAIINPTVMSY 300  
 20 GSGNHCHNMLFDN+GNNAFF+D R GMQLSEDAITFLGLMKHAYNTAI NPTVMSY 300  
 Sbjct: 241 AGSGNHCHNMLFDNCGNNAFFYDEADKRGMLSEDAITFLGLMKHAYNTAITNPTVMSY 300

Query: 301 KRLVPGYEAPVYVAVAGRNRSPLIRVPASRGMTKLELRSVDPTANFYAL+VLL +GL+ 360  
 25 KRLVPGYEAPVYVAVAG NRSPILIRVPASRGMTKLELRSVDPTANFYAL+VLL +GL+ 360  
 Sbjct: 301 KRLVPGYEAPVYVAVAGSNRSPLIRVPASRGMTKLELRSVDPTANFYALALVLEGLD 360

Query: 361 GIENKIBAPEPISTINIAWYVEERRQAGIVDLPSLTHNALALAEDEVVKAALGTHIYTN 420  
 30 GI NKBAPPE+E NIY NT-EER +AGI+DLPSLTHNAL+AL+D+VV+ ALG HITTN 420  
 Sbjct: 361 GIENKIBAPEPISTINIAWYVEERRNEAGIIDLPSLTHNALALQDDVVQKALGYHIYTN 420

Query: 421 FLDAKRIEWASAYTYSQWEIDNLYLY 448  
 FL+AKRIEW+SYAT+VSQWEID+Y+ Y 448  
 Sbjct: 421 FLBAKRIEWSAYTYSQWEIDHYIYNY 448

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1586

- A DNA sequence (GBSx1680) was identified in *S. agalactiae* <SEQ ID 4903> which encodes the amino acid sequence <SEQ ID 4904>. This protein is predicted to be SceB precursor. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 50 >GF:CAA66624 (GB: X97985) ORF1 [Staphylococcus aureus]  
 Identities = 44/119 (36%), Positives = 66/119 (54%), Gaps = 4/119 (3%)
- Query: 26 SFASTNADANTYNYAVVDYLASAEITQAHPA-SHTFPLQCTGWGKE-MATVAGNMG 83  
 55 S AS + +N + ++ I+ + + GN + GQCT+ V + + G+ WG 176  
 Sbjct: 117 SGASYSTTENNIVHVTITAPSGNRSISISNYASGNLYTSQCTYIVFDRVGGIKGSGTWG 176
- Query: 84 NGSDKWAASASADYTVGTQPRVSGIYVWITGSGYGHVAYTVAVDPVINKIQVLEENYAGH 142  
 N +WA +AAS+ YTV P-VG+I+ T G YGHVAYV V+ ++V E NY GH 142  
 Sbjct: 177 NASNWNANAAASGGYTVNNTPKVGRIMTITQGYGHVAYVSGVNS-NGSVRVSESNY-GH 233
- 60

-1772-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1013> which encodes the amino acid sequence <SEQ ID 1014>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 60/115 (52%), Positives = 81/115 (70%), Gaps = 7/115 (6%)

Query: 55 AHPASNTFFPLGQCTWGVKEMATWAGNWWNGGGDWAAASADYTVGTQPRVGSIVCVTDG 114
      ++ +SNT+P+GQCTW G K +A WAGN WNGG WR SA +A Y G+ P VG+I W DG
Sbjct: 291 SYDSNTYFVGQCTWGAKSAPWAGNWWNGGGQWYSAQAAGTGTGTPMVGLAVWDG 350

Query: 115 SYGHVAVYTAVDVPVINKIQLSNTAGHKNIDITGNFQFNITVTPGVVSYIYEN 169
      YGHVA V V ++ I+V+ESNI+G Q+I ++RGNF+P V++IYP+
Sbjct: 351 GYGHVAVVVEVQSSG-IRVWESNYSGRQYIADIRGNFNPFG-----VTFTIYPH 398

```

A related GBS gene <SEQ ID 8859> and protein <SEQ ID 8860> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: 5.85
GVH: Signal Score (-7.5): 3.11
      Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 6.74 threshold: 0.0
PERIPHERAL Likelihood = 6.74 115
modified ALOM score: -1.85

*** Reasoning Step: 3

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

37.5/56.7% over 200aa
GP|1340128| ORF1 Insert characterized Staphylococcus aureus

ORF00255(376 - 726 of 1107)
GP|1340128|emb|CAA6624.1||X97985(33 - 233 of 255) ORF1 {Staphylococcus aureus}
%Match = 9.0
%Identity = 37.5 %Similarity = 56.7
Matches = 45 Mismatches = 47 Conservative Sub.s = 23

294 324 354 384 414
SVLWI+*TRSHQKRENMMNIKQKSKTMLGTVALVSAFSAFSTNADANNTYNYAVDVD-----
      :
55 MKKIVATITATAGLATIATFAGHIDAQAARQNNNGVYNSNDAGSYSEYTYTIDAQSNHYHYTWGNVNSQLQNN-----
      10 20 30 40 50 60 70

462 489 516 546 576 606
-----YLASAKRIQAHPA-SNTFFPLGQCTWGV-KEMATWAGNWWNGGGDWAAASADYTVGTQ
      :
60 GSGASYSITSNVHVTTAAPSNGKRSISNGYASGSNLYTSGQCTYYVFDREVGGKIGSTVGNASNVAAPASGGYTVNMT
      130 140 150 160 170 180 190

```

-1773-

```

636      666      696      726      756      786      816      846
PRVGSIVCWDSSYGHVAVYVAVDPVTNKGVLSENKAGHWNIDNYRGWEDPQNTVTPGVVSYIYFN*SIGNSHRRYKS
|:|:|: | | ||||| | : :| | | |
PKVGAIMQTTCGYGHHVAVYVGVNS-NGSVRVSENNY-GHGAGVVTSTRTISANQAGSYNFIH
      210      220      230      240      250

```

SEQ ID 8860 (GBS30) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 2; MW 19.2kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 2; MW 44.2kDa).

GBS30-GST was purified as shown in Figure 193, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1587

A DNA sequence (GBSx1681) was identified in *S. agalactiae* <SEQ ID 4905> which encodes the amino acid sequence <SEQ ID 4906>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood = -3.93    Transmembrane    2 - 18 ( 1 - 18)
----- Final Results -----
bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1588

A DNA sequence (GBSx1682) was identified in *S. agalactiae* <SEQ ID 4907> which encodes the amino acid sequence <SEQ ID 4908>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2160(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 353/550 (64%), Positives = 443/550 (80%)

Query: 6 LKPREVGVYAIGGLGEIGKNTFYGIEYQDEILIVDAGIKFPEDLLGIDYVIPDYSYIVEN 65
LK + VYA+GGLGEIGKNTFY +++QDEII++DAGIKFPED+LGIDYVIPDYSY+V+N
Sbjct: 4 LKNNQTAVALGGLGEIGKNTFYAVQDEILIDAG:KFPEDLLGIDYVIPDYSYLVN 63

Query: 66 IDRIKALVITHGHEDHIGGIPFLLQANLPTIYAGPLAALAIKGLRRHGLLRDATLYEIR 125
++IK L ITHGHEDHIGGIP+LL++ N+PIY G LAL L++GKLRHGLLR A L++I
Sbjct: 64 ENKIKGLPITHGHEDHIGGIPYLLRRVNIPIYGGKIALGLLRGKLRHGLLRKAKLRIDQ 123

```

-1774-

Query: 126 ANTELTPKNLSVTFFRTTHSIPEPLGIVIHPTQGVICTGDFKFDFTPVGEPADLRHMA 195  
 + F SV+FFRTTHSIP+ GIV+ TP G++ TGDFKFDFTPVGEP+L+MA  
 5 Sbjet: 124 EDDIIFKAKTSVSFFRTTHSIDPSYGIIVKTPPGNIVHTGDFKFDFTPVGEPAMIKMAK 183

Query: 186 LGSEGVLCCLSDSTNAEIVPTFTNSEKIVQSIMKIIISGIGRIIPASPASNIIRLQQA 245  
 +GE+GVLCCLSDSTN+E+P FT SE+ VG+SI I +BGRIP+PASNI RLQQA E  
 Sbjet: 184 IGSEGVLCCLSDSTNSEIPEFTMSERKVGESIDHIFPRVGRRIIPATFASNIHRLQQA 243

10 Query: 246 AAVTKGRKIAVFGSRMRKIAVINGIKLVYIKVPKGTPIKPSKLNHASEVLIMCTGSQGE 305  
 +AV+ GRK+AVFGSRME AI G ELGVYK PK FTIEP++L L +EV+I+CTGSQGE  
 Sbjet: 244 SAVRYGRKAVFGRSMESAINIGQELGYIKAPKNTFTFNPQLNKLIPNEVMILCTGSQGE 303

15 Query: 306 SMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNITSVNKLINTIQEAGVDVINGKIMNI 365  
 MAAL+R+R+ GTHRQ+ + PGDTVIFSSSPIPGNT SV+K IN + +AG +VIHG +N+I  
 Sbjet: 304 PMAALSRVAPGTHRQICIIIGDTVIFSSSPIPGNTLSVSKITINQLYKAGANVINGSLNDI 363

20 Query: 366 HTSGHGQQSQKLMRLIKPKYFMPVHGKYMQRVHAGLAVDTGPKENITFIMENGVDLA 425  
 HTSGHGQQ+SQKLMRLIKPKYFMP+HGEYRM K+H LA D G+P EN FIM+NGVDLA  
 Sbjet: 364 HTSGHGQQSQKLMRLIKPKYFMPHIGYRMLKMHMTKLAEDGVPAENCFIMENGVDLA 423

25 Query: 426 LTSDSARIAGHPNAQDIYVDGNGIGDGAUVRDRHDLSDGVLVAATVDFDSRIMLAG 485  
 L D A IAG + +YVDGNGIGDIG VLDRD LSE+G+V+ V +++ + AG  
 Sbjet: 424 LHPDEAGTAGIKPSGVTVYVDGNGIGDIGNVLDRRIILSEGLVNVVVVSLNMKRYKVTAG 483

Query: 486 PDILSRGFVYMRSGDLIRSECHILFNATRIALKNDASIQVNGAIVNALRPFLYEKTE 545  
 PD++SRGF+YMRSGDLI+E+Q +L N ++ ++ K + I + L PFLY++T+  
 Sbjet: 484 PDILSRGFVYMRSGDLIQEACRLLANHLQSVMERKTNQWSEIKNEITDVLGPFLYDRTK 543

30 Query: 546 REPIIIPMWL 555  
 R+P+I+P+++  
 Sbjet: 544 RKMILPIIM 553

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4909> which encodes the amino acid  
 35 sequence <SEQ ID 4910>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.11 Transmembrane 468 - 484 ( 468 - 484)

40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1044 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]  
 Identities = 353/550 (64%), Positives = 444/550 (80%)

50 Query: 6 LKPNVGVFPAIGLGLGKNTGYGIEYQDRIIIVDAGIKFPEDDLGLSDIVIPDYSYVDN 65  
 LK N+ V+R+GGEGRIEGRNTY +++QDEII+DAGIKFPED+LLGLSDIVIPDYSY+V N  
 Sbjet: 4 LKNNQAVYALGGLGKNTGYAVQDQRIILIDAGIKFPEDDELGLSDIVIPDYSYLVKN 63

55 Query: 66 LDRVKALVITHGHEDHIGGIPFLKQANITPYAGPLAIALIRGLEHGLMRKATVYEIN 125  
 +++K L ITHGHEDHIGGIP+LL++ NIPY G LAL L+RGKLEHGL R+A +++I  
 Sbjet: 64 ENKIKGLFTYTHGHEDHIGGIPYLLREVNIPYGGKIALGLLRKLEHGLLRKAKLHDIQ 123

Query: 126 HNTLTFKMSVTFPFTTHSIPRPVGIHPTQSKIICGDFKFDFTPVGEPADLRHMA 185  
 + F SV+FF+TTHSIP+ GIV+ TP G I+ TGDFKFDFTPVGEP+L+MA  
 60 Sbjet: 124 EDDIIFKAKTSVSFFRTTHSIDPSYGIIVKTPPGNIVHTGDFKFDFTPVGEPAMIKMAK 183

Query: 186 LGSEGVLCCLSDSTNAEIVPTFTNSEKIVQSIMKIIISGIGRIIPASPASNIYRLQQA 245  
 +GE+GVLCCLSDSTN+RIP FT SE+ VG+SI I +GRIP+PASNI+RLQQA E  
 Sbjet: 184 IGSEGVLCCLSDSTNSEIPEFTMSERKVGESIDHIFPRVGRRIIPATFASNIHRLQQA 243

65



-1775-

Query: 246 AAVKTRKIAVFGSRMEKAIVNGILGYIKVPGKTFIEPSELKNLHASEVLIMCTGSQGE 305  
 +AV+ GRK+AVFGSRME AI G ELGYIK PK TFIEP+L L +EV+I+CTGSQGE  
 Sbjct: 244 SAURYGRKIAVFGSRMESAINIGELGYIKAPKNTFIEPWLNKL PNEVMILCTGSQGE 303

5 Query: 306 SMAALARIANGTHRQVTLPQGDVTFSSSPIKGNITSVNKLINTIQEAGVDVIHGKVNNI 365  
 MAAL+R+A GTHRQ+ + PGDTVFSSSPIKGNIT SV+K IN + +AG +VIHG +N+I  
 Sbjct: 304 PMAALSRVAGTHRQVTLPQGDVTFSSSPIKGNITLSVSKTINQLYKAGANVHGSLNDI 363

10 Query: 366 HTSGHGGQQBQKMLMLIKPKYFMPVHGEYRMQKVHAGLMDIGIPKENIFIMENGDLVA 425  
 HTSGHGGQ+BQKMLL LIKPKYFMP+HGEYRM K+H LA D G+P EN FIM+NGDLVA  
 Sbjct: 364 HTSGHGGQBQKMLMLIKPKYFMP LHGEYRMKRMHTKLARDQGVPAKNCFTMDNGDLVA 423

Query: 426 LITSDSARIAGHPNAQDIYVDGNGIGDIGAVALDRDRDLSDGVLVAVATVDFNTQIMILAG 485  
 L D A TAG + +YVDGNGIGDIG VLRDR LSE+G+V+ V +++ + AG  
 15 Sbjct: 424 LHFEAGIAGIKPGSVYVDGNGIGDIGNIVLRDRRLSEBGLVVVVVSLMKKEYKVLAG 483

Query: 486 PDILSRGFYIMRESGDLIRESQVLFPNAIRIALKNKDISQSVNGAIVNALRPFLYEKTE 545  
 PD+SRGF+YIMRESGDLI+R+QR+L N ++ ++ K + I + L PFLY++T+  
 20 Sbjct: 484 PDILSRGFYIMRESGDLIQEALRLLANHLQEVMERKTNQWSEIKNEITDVLGPFYDRTK 543

Query: 546 RRFIIIPMVL 555  
 R+P+I+P+++  
 Sbjct: 544 RKFMLPIIM 553

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 523/559 (93%), Positives = 550/559 (97%)

Query: 1 MSNINLKPEZGVYAIGGLGEIGKNTYIGIEYQDEIIIVDAGIKFPEDLLGIDYVYPDS 60  
 K+NI+LKP EVGV+AIIGLGEIGKNTYIGIEYQDEIIIVDAGIKFPEDLLGIDYVYPDS  
 30 Sbjct: 1 MTNISLKPNEVGVAIGLGEIGKNTYIGIEYQDEIIIVDAGIKFPEDLLGIDYVYPDS 60

Query: 61 YIVNITRIKALVITGHEDHIGGIPFLKQANLPIYAGPLALAIKLGIEBHLKRLDRTAT 120  
 YIV+N+DR+KALVITGHEDHIGGIPFLKQAN+PIYAGPLALALI+KLGIEBHL R+AT  
 Sbjct: 61 YIVNIDRVLKALVITGHEDHIGGIPFLKQANLPIYAGPLALALIRGLIEBHLMRERAT 120

35 Query: 121 LYRIHANTLTFQNLVTFTRTHSIPEPLGIVINTPQGRVICTGDPKDFPTVPGEPADL 180  
 +YBI+ NTELTFQN+SVTFK+TTHSIPEP+GIVINTPQGR+ICTGDPKDFPTVPV+PADL  
 Sbjct: 121 VYIHNNTLTFQNSVTFTRTHSIPEPVGIVINTPQGI ICTGDPKDFPTVPVGPADL 180

40 Query: 181 HRNALGEGVLCLLSDSTNAEVPPTFNSKIVGQSIMKITEGIBGRIIFASFASNI+RL 240  
 RNAALGE+GVLCLLSDSTNAE+PTFNTSEK+VQSI+KIEGI GRIFASFASNI+RL  
 Sbjct: 181 RNAALGEGVLCLLSDSTNAEIPPTFNTSEKIVGQSIKITEGIBGRIIFASFASNIYRL 240

Query: 241 QQAEEAAVKTGRKIAVFGSRMEKAIVNGILGYIKVPGKTFIEPSELKNLHASEVLIMCT 300  
 QQAEEAAVKTGRKIAVFGSRMEKAIVNGILGYIKVPGKTFIEPSELKNLHASEVLIMCT  
 45 Sbjct: 241 QQAEEAAVKTGRKIAVFGSRMEKAIVNGILGYIKVPGKTFIEPSELKNLHASEVLIMCT 300

Query: 301 GSQGSMAALARIANGTHRQVTLPQGDVTFSSSPIKGNITSVNKLINTIQEAGVDVIHG 360  
 GSQGSMAALARIANGTHRQVTLPQGDVTFSSSPIKGNITSVNKLINTIQEAGVDVIHG  
 50 Sbjct: 301 GSQGSMAALARIANGTHRQVTLPQGDVTFSSSPIKGNITSVNKLINTIQEAGVDVIHG 360

Query: 361 KINNITHTSGHGGQQBQKMLMLIKPKYFMPVHGEYRMQKVHAGLMDIGIPKENIFIMEN 420  
 K+NNITHTSGHGGQQBQKMLL LIKPKYFMPVHGEYRMQKVHAGLMDIGIPKENIFIMEN  
 Sbjct: 361 KVNNTHTSGHGGQQBQKMLMLIKPKYFMPVHGEYRMQKVHAGLMDIGIPKENIFIMEN 420

55 Query: 421 GDVIALTSDSARIAGHPNAQDIYVDGNGIGDIGAVALDRDRDLSDGVLVAVATVDFNSK 480  
 GDVIALTSDSARIAGHPNAQDIYVDGNGIGDIGAVALDR DLSDGVLVAVATVDF+++  
 Sbjct: 421 GDVIALTSDSARIAGHPNAQDIYVDGNGIGDIGAVALDRDRDLSDGVLVAVATVDFNTQ 480

60 Query: 481 MILAGPDILSRGFYIMRESGDLIRESQHLFPNAIRIALKNKDISQSVNGAIVNALRPFL 540  
 MILAGPDILSRGFYIMRESGDLIRESQ +LFPNAIRIALKNKDISQSVNGAIVNALRPFL  
 Sbjct: 481 MILAGPDILSRGFYIMRESGDLIRESQVLFPNAIRIALKNKDISQSVNGAIVNALRPFL 540

Query: 541 YEKTEREPIIIPMVLTPDK 559  
 YEKTEREPIIIPMVLTPDK  
 65 Sbjct: 541 YEKTEREPIIIPMVLTPDK 559

-1776-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1589**

A DNA sequence (GBSx1683) was identified in *S. agalactiae* <SEQ ID 4911> which encodes the amino acid sequence <SEQ ID 4912>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2932 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAH13327 GB:Z99111 ykzG [Bacillus subtilis]
Identities = 27/75 (36%), Positives = 44/75 (58%), Gaps = 7/75 (9%)

Query: 1 MIYKVFYQETKERNPRREQTKTLTYVTIDAANLESGRIAARKLVVENTAYNIEFIELLSK 60
        MIYKVFYQGE + P RE+T +LY+ + ++ ++ +K +NIEFI +
Sbjct: 1 MIYKVFYQGEKAEDEVVREKIDSLYIEGVSEKDVRTKLKEKK-----FNISFITPVDSA 53

Query: 61 HLEYEKETGVFELTE 75
        LEYE+++ F++ E
Sbjct: 54 FLEYEQSQENFKVLS 68

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4913> which encodes the amino acid sequence <SEQ ID 4914>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3428 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 60/76 (78%), Positives = 70/76 (91%)

Query: 1 MIYKVFYQETKERNPRREQTKTLTYVTIDAANLESGRIAARKLVVENTAYNIEFIELLSK 60
        MIYKVFYQETK+++PRRE TK LY+ IDA +EL+GRI AR+LVE+NT YN+EFIELLSK
Sbjct: 1 MIYKVFYQETKQSPREESTKALYLNIDATDELGDRIKARRLVEDNTYNNYFIELLSK 60

Query: 61 HLEYEKETGVFELTEP 76
        HL+YEKSGVFELEP
Sbjct: 61 HLEYEKSGVFELEP 76

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1590**

A DNA sequence (GBSx1684) was identified in *S. agalactiae* <SEQ ID 4915> which encodes the amino acid sequence <SEQ ID 4916>. This protein is predicted to be glycoprotein endopeptidase. Analysis of this protein sequence reveals the following:

```

Possible site: 13

```

-1777-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence (or aa 1-17)

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.0430(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA76861 GB:Y17797 hypothetical protein [Enterococcus faecalis]  
 Identities = 94/182 (51%), Positives = 127/182 (69%), Gaps = 6/182 (3%)

Query: 2   MKVLAFTDTSKALSVAVLNMECLATVTINIKKNHSINLMPALDFLMQSIDLEPQDLRI 61  
           +++LA DTS++ LS+AV N + L + T +K+NHS+ LMPALD+LM ++L P +DR  
 Sbjct: 13 VRLLAIDTSNQTLSIACVENCQKILGSTATVKNHSLTLMFALDYLMSQLNLAPTAIDRF 72

15 Query: 62 VVABGPGSYTGLRVAVATAKMLAYTLKIDLGVSSLYAL-TNPFSENDILVPLIDARENN 120  
           VVABGPGSYTGLR+ V TAK IAYTLK +LVG+SSL AL N + L+VPL DNR N  
 Sbjct: 73 VVABGPGSYTGLRGLVITAKTLAYTLKELVGISSQLALANCVGQTGLIVLPLDFDARRKN 132

20 Query: 121 VYVGPFYQNGDITV----KPDCHTSLEEVLEQVGKNKANHVFGE-VAAFPDQIKKALPHAKI 175  
           VY G Y+ D V PD H SL E+I+++ N+ N+ FVGE V F ++I + +PH +I  
 Sbjct: 133 VYAGAYRFVDGVWGNELPDQHISLRLLBQLKNEPNLFFVGEDVEKFTETIAQIHFGHEI 192

Query: 176 TE 177  
 Sbjct: 193 CD 194

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4917> which encodes the amino acid sequence <SEQ ID 4918>. Analysis of this protein sequence reveals the following:

30 Possible site: 36  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.38 Transmembrane 99 - 115 ( 99 - 115)

35 ----- Final Results -----  
           bacterial membrane --- Certainty=0.1553(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9159> which encodes the amino acid sequence <SEQ ID 9160>. Analysis of this protein sequence reveals the following:

40 Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.38 Transmembrane 88 - 104 ( 88 - 104)

45 ----- Final Results -----  
           bacterial membrane --- Certainty=0.1553(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 134/232 (57%), Positives = 172/232 (73%), Gaps = 3/232 (1%)

Query: 2   MKVLAFTDTSKALSVAVLNMECLATVTINIKKNHSINLMPALDFLMQSIDLEPQDLRI 61  
           MK LAFDTS+K LS+AA+L+ LA +T+NI+K HS+LMPALDFLM DL+PQDL+RI  
 Sbjct: 12 MKTLAFTDSNKILSLALLDDESLADNTLNIQKHSVSLMFAIDFLMCTDLKPOQLERI 71

55 Query: 62 VVABGPGSYTGLRVAVATAKMLAYTLKIDLGVSSLYALTNGFSE---NOLLVPLIDARR 118  
           VVA+GPGSYTGLRVAVATAK LAY+L I LVG+SSLAL + N L+VPLIDARR  
 Sbjct: 72 VVABGPGSYTGLRVAVATAKFLAYSINLALVVISLALAASTCKQYPTILVPLIDARR 131

60 Query: 119 NNVYVGPFYQNGDITV-KPDCHTSLEEVLEQVGKNKANHVFGEVAAFPDQIKKALPHAKITET 178  
           N YVG+Y+ G +V P H SL E +++++ + + FVGE A F ++I+K LP A + T

-1778-

Sbjct: 132 QNAYVGYRQGSVMPQAHSLRVIIIEQLVBEQQLIFVGTAPPAKKIQKKLPQAILLEF 191

Query: 179 LPCAVATGRGQKQKSVNVDAFVPRYLKRVABENWLNHCETNTEBYIKRV 230  
 LP A G GQ + NVDAFVP+YLKRVABENWLNK++ + Y+KR+

Sbjct: 192 LPSAYCEGLLGQSLAPENVDAFVPQYLKRVABENWLNKDNKIDDSHYVERK 243

SEQ ID 4916 (GBS69) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 9; MW 28.9kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 4; MW 53.9kDa).

The GBS69-GST fusion product was purified (Figure 197, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 285), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

# Example 1591

A DNA sequence (GBSx1685) was identified in *S. agalactiae* <SEQ ID 4919> which encodes the amino acid sequence <SEQ ID 4920>. This protein is predicted to be ribosomal-protein-alanine acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10137> which encodes amino acid sequence <SEQ ID 10138> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA006803 GB:AB000696 ribosomal-protein-alanine acetyltransferase  
 [Aqifex aeolicus]

Identities = 44/141 (31%), Positives = 74/141 (52%), Gaps = 8/141 (5%)

Query: 9 LRFEMESBQALAINSVLSGVVDKSPWLSQISEDLKKDSTDYFFVYNDGEVIGFLALQ 68

+RE E E E+ ++ + + WS +D + + P+ DG+V+G++

Sbjct: 4 VRMEREDVER---VYEINRESPTTDWRSRPSFEKDFENKPSRFLVEDGKVVGVIFW 60

Query: 69 GLVGEVEITNIAVKKNYQGRGAYQIM----SMIADLEVPPVFLVRYSNIVAQKLNERCG 124

+ E I A+ Y+GKG +L+ S + D V L+VR SN+ A LY+ G

Sbjct: 61 VVKEBATIMFALAPGYRGKYGKELREATSRLGDKVRGVVIVRKNRLAINLYKILG 120

Query: 125 FVVLKRFKNFYHDPIKDAIWM 145

F V+ +HK YY D E+A++M

Sbjct: 121 FKVVTERGGYSDG-ENALLM 140

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4921> which encodes the amino acid sequence <SEQ ID 4922>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3800 (Affirmative) < succ>

-1779-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5       Identities = 65/140 (46%), Positives = 96/140 (68%), Gaps = 1/140 (0%)

Query: 9   LREFEMES-SEQALAINSVLSVDYDKSPWLSQISELKKDSTDYFFVYNDGEVIGFLAL 67  
           L E M++ EQA I+ +L VY SPW+L Q+ D++D TDYF +Y+ ++GFLA+

10       Sbjct: 6   LSSNMKTVERQAKNTYQLLENVYGTSPWTLSEQVLIDIRRDQTDYFLLYDHDKLLGFLAI 65

Query: 68   QQIAGSEVRIYTNIAVKKNYQGGKYAYQLMSMIADIEVPVFLSVRYSNIVAQKLYERGCFV 127  
           Q L GEVE+T IA+ ++Q G A QLM+ + IE +FLSVR SN AQ LY++ GF

          Sbjct: 66   QDLAGEVPMQTIALPQHQLGLASQLMTHLDSIESDIFLEVRSSNHRACGLYQKPKFK 125

15       Query: 128   LAKRKNYYHDPIEDAIVMRK 147  
                   + KR +YY +PIE A++M++

          Sbjct: 126   IGKRPDYRNPDIETALLMRK 145

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 20   vaccines or diagnostics.

#### Example 1592

A DNA sequence (GBSx1686) was identified in *S.agalactiae* <SEQ ID 4923> which encodes the amino acid sequence <SEQ ID 4924>. Analysis of this protein sequence reveals the following:

25       Possible site: 21  
        >>> Seems to have no N-terminal signal sequence

       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 35   vaccines or diagnostics.

#### Example 1593

A DNA sequence (GBSx1687) was identified in *S.agalactiae* <SEQ ID 4925> which encodes the amino acid sequence <SEQ ID 4926>. Analysis of this protein sequence reveals the following:

40       Possible site: 38  
        >>> Seems to have no N-terminal signal sequence

          INTEGRAL   Likelihood = -1.75   Transmembrane   86 - 102 ( 86 - 104)

       ----- Final Results -----  
           bacterial membrane --- Certainty=0.1702 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45

The protein has homology with the following sequences in the GENPEPT database.

50       >GP:BAB04267 GB:AP001508 glycoprotein endopeptidase [Bacillus halodurans]  
        Identities = 194/331 (58%), Positives = 263/331 (78%), Gaps = 1/331 (0%)

       Query: 6    ILAVSSCDSTSVAILKNDKELLANLIASQVSHKRRGGVVPEVASRHVVEVVTTCFEDA 65  
                   ILA+E+SCDST A++N   +L+N+++SQ++SHKRRGGVVPE+ASRHVE   T   E+A

-1780-

Sbjct: 12 ILAIFTS~~CD~~ETSA~~AV~~IE~~NG~~TTILSNVVSQIDSHK~~RG~~VV~~VE~~IAS~~RH~~VE~~QIT~~IV~~VE~~A 71

Query: 66 LQ~~EG~~IAS~~DL~~DA~~VA~~VTG~~PL~~VG~~AL~~L~~V~~GA~~AA~~KA~~FA~~W~~N~~KL~~EL~~IP~~IN~~IM~~AG~~H~~MA~~ARDV 125  
+ EAG+ +DL AVAVT G~~PL~~VG~~AL~~L+G+ AA~~KA~~ A+A++LELI ++H+AGH+ A R +

5 Sbjct: 72 MHEAGVDFAD~~LA~~AV~~VT~~SG~~FL~~VG~~AL~~L~~I~~GV~~NA~~AKA~~IA~~FA~~HC~~LE~~LI~~GV~~HH~~AG~~HI~~YAN~~RL~~L 131

Query: 126 KELQY~~PL~~L~~ALL~~V~~SG~~HT~~EL~~V~~Y~~VE~~SP~~GDY~~KI~~V~~GE~~TRDDA~~VG~~E~~AN~~DKV~~RG~~V~~GL~~T~~YP~~AG~~RE~~I 185  
KEL++PL~~AL~~-V~~SG~~HT~~EL~~+Y+ G+++++GET~~RD~~DA~~VG~~E~~AN~~DKV R +GL YP G I

10 Sbjct: 132 KELE~~FF~~PL~~AL~~V~~SG~~HT~~EL~~IY~~ME~~NH~~GE~~FE~~VI~~GET~~RD~~DA~~VG~~E~~AN~~DKV~~RT~~LG~~LP~~Y~~GG~~PH~~I~~ 191

Query: 186 DQLAH~~KG~~Q~~DT~~Y~~HP~~FRAM~~IK~~DE~~HL~~FS~~FG~~LK~~SA~~FIN~~I~~HN~~AK~~Q~~KG~~E~~AL~~V~~ED~~L~~CA~~SP~~QA~~ 245  
D+LA G+IT F~~PR~~A++D +F~~SF~~SG~~LK~~SA IN HNA+Q+GE + ED+ AS~~FQ~~A+

15 Sbjct: 192 DRLAV~~NG~~ED~~TL~~Q~~FR~~AW~~LE~~FD~~S~~DF~~SF~~SG~~LK~~SAV~~INT~~L~~HN~~AK~~Q~~RG~~EN~~V~~Q~~AE~~DA~~V~~AS~~F~~QA~~S 251

Query: 246 VLDI~~LL~~ANT~~Q~~K~~AL~~L~~K~~Y~~P~~FK~~TL~~V~~AG~~GV~~AA~~N~~Q~~GL~~ER~~LA~~TD~~IS~~PD~~-ID~~VP~~IP~~PL~~AL~~CG~~DA 304  
V+D+L+ KT+KA +Y V+ ++AGG~~VA~~AN+GLR L + ID+V~~IP~~PL LC DNA

20 Sbjct: 252 VID~~LV~~IT~~TK~~KA~~AB~~EY~~KV~~RQ~~V~~LL~~AG~~GV~~AA~~N~~GL~~RT~~AL~~EE~~AF~~FE~~PI~~DL~~VI~~P~~PL~~SL~~CT~~DA 311

Query: 305 GMIAL~~AA~~A~~IE~~FE~~KN~~FAS~~LK~~IM~~AK~~PS~~LA~~FES 335  
MI AA+I+F+++ QA + LN +PSL E+

25 Sbjct: 312 AMIG~~AA~~AS~~IK~~FK~~QT~~FA~~GM~~DL~~NK~~Q~~PS~~LE~~LN~~ 342

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4927> which encodes the amino acid sequence <SEQ ID 4928>. Analysis of this protein sequence reveals the following:

25 Possible site: 38  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -2.76 Transmembrane 86 - 102 ( 85 - 104)

30 ----- Final Results -----  
bacterial membrane --- Certainty=0.2105 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:EAB04267 GB:AP001508 glycoprotein endopeptidase [Bacillus halodurans]  
Identities = 196/330 (59%), Positives = 255/330 (76%), Gaps = 2/330 (0%)

Query: 6 ILAV~~ES~~SC~~DE~~TS~~AI~~LK~~NE~~ST~~LL~~SN~~VI~~AS~~QV~~ES~~HK~~RG~~GV~~V~~VE~~IAS~~RH~~VE~~QIT~~IV~~VE~~A 65  
ILA+E+SC~~DE~~TS A+++N +T+L~~SN~~V++SQ++SHK~~RG~~VV~~VE~~IAS~~RH~~VE~~QIT~~IV~~VE~~A

40 Sbjct: 12 ILAIFTS~~CD~~ETSA~~AV~~IE~~NG~~TTILSNVVSQIDSHK~~RG~~VV~~VE~~IAS~~RH~~VE~~QIT~~IV~~VE~~A 71

Query: 66 LQ~~EG~~IAS~~DL~~DA~~VA~~VTG~~PL~~VG~~AL~~L~~V~~GA~~AA~~KA~~FA~~W~~N~~KL~~EL~~IP~~IN~~IM~~AG~~H~~MA~~ARDV 125  
+ EAG+ +DL AVAVT G~~PL~~VG~~AL~~L+G+ AA~~KA~~ A+A+ LELI V+H+AGH+ A R

45 Sbjct: 72 MHEAGVDFAD~~LA~~AV~~VT~~SG~~FL~~VG~~AL~~L~~I~~GV~~NA~~AKA~~IA~~FA~~HC~~LE~~LI~~GV~~HH~~AG~~HI~~YAN~~RL~~L 131

Query: 126 KPL~~AV~~PL~~IAL~~L~~V~~SG~~HT~~EL~~V~~Y~~VE~~SP~~GD~~Y~~KI~~V~~GE~~TRDDA~~VG~~E~~AN~~DKV~~RG~~V~~GL~~T~~YP~~AG~~RE~~I 185  
K L +PL+AL-V~~SG~~HT~~EL~~+Y+ G++ +IGET~~RD~~DA~~VG~~E~~AN~~DKV R +GL YP G I

50 Sbjct: 132 KELE~~FF~~PL~~AL~~V~~SG~~HT~~EL~~IY~~ME~~NH~~GE~~FE~~VI~~GET~~RD~~DA~~VG~~E~~AN~~DKV~~RT~~LG~~LP~~Y~~GG~~PH~~I~~ 191

Query: 186 DQLAH~~KG~~Q~~DT~~Y~~HP~~FRAM~~IT~~ED~~HL~~FS~~FG~~LK~~SA~~FIN~~I~~HN~~AK~~Q~~KG~~E~~AL~~V~~ED~~L~~CA~~SP~~QA~~ 245  
D+LA G+IT F~~PR~~A+ D +F~~SF~~SG~~LK~~SA IN HNA~~K~~G+G+ + ED+ AS~~FQ~~A+

55 Sbjct: 192 DRLAV~~NG~~ED~~TL~~Q~~FR~~AW~~LE~~FD~~S~~DF~~SF~~SG~~LK~~SAV~~INT~~L~~HN~~AK~~Q~~RG~~EN~~V~~Q~~AE~~DA~~V~~AS~~F~~QA~~S 251

Query: 246 VLDI~~LL~~ANT~~Q~~K~~AL~~S~~Y~~PA~~NK~~MA~~VA~~GV~~AA~~N~~Q~~GL~~ER~~LA~~Q~~E~~II~~--TH~~IV~~IP~~PL~~AL~~CG~~DA 303  
V+D+L+ KTKKA Y + ++AGG~~VA~~AN+GLR L + I+V~~IP~~PL LC DNA

60 Sbjct: 252 VI~~NL~~VI~~TK~~KA~~AB~~EY~~KV~~RQ~~V~~LL~~AG~~GV~~AA~~N~~GL~~RT~~AL~~EE~~AF~~FE~~PI~~DL~~VI~~P~~PL~~SL~~CT~~DA 311

Query: 304 GMIAL~~AA~~A~~IE~~Y~~KD~~HFAN~~SM~~IM~~AK~~PS~~LA~~FD 333  
MI AA+I++ +Q FA M LN +PSL +

Sbjct: 312 AMIG~~AA~~AS~~IK~~FK~~QT~~FA~~GM~~DL~~NK~~Q~~PS~~LE~~LN~~ 341

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/334 (86%), Positives = 313/334 (93%), Gaps = 1/334 (0%)

-1781-

Query: 1 MKDRYILAVSSCDRTSVAILLKNDKRLLANI L ASQVESHRRFGGVVPEVASRHHVEVVT 60  
 M DRYILAVSSCDRTSVAILLKNDKRLLANI L L+N+ L ASQVESHRRFGGVVPEVASRHHVEV+TT  
 Sbjct: 1 MDRYILAVSSCDRTSVAILLKNDKRLLANI L ASQVESHRRFGGVVPEVASRHHVEVVT 60

5 Query: 61 CFPDQLQKAGIVASDLQAVVTVYGGIAGALLVGMMAAKAFAMANKLPLIPINHMAGHLM 120  
 CFPDQLQKAGI ASDI L AVAVTVYGGIAGALLVGMMAAKAFAMANKLPLIP+NHMAGHLM  
 Sbjct: 61 CFPDQLQKAGI SASDLSAVAVTVYGGIAGALLVGMMAAKAFAMANKLPLIPVNHMAGHLM 120

10 Query: 121 AARDVKELQYPLIALLVSGGHTSLVYVSEPGDYKIVGSTRDDAVGEAYDKVGRVMGLTYP 180  
 AAR+ K L YPL+ALLVSGGHTSLVYV EPDGY I+GETRDDAVGEAYDKVGRVMGLTYP  
 Sbjct: 121 AARQKPLVYPLIALLVSGGHTSLVYVSEPGDYHITIGSTRDDAVGEAYDKVGRVMGLTYP 180

15 Query: 181 AGREIDQLAHKGQUTYHFFRAMIKEDHLEFSPGLKSAFINLHRNAQKQKALVLEDCAL 240  
 AGREIDQLAHKGQUTYHFFRAMI EDHLEFSPGLKSAFINLHRNA+KQG+ L+LEDLCA  
 Sbjct: 181 AGREIDQLAHKGQUTYHFFRAMITEDHLEFSPGLKSAFINLHRNAQKQKALVLEDCAL 240

20 Query: 241 SFQRAVLDDLLAKTKKALLKYPVKTLVAVGGVANNQGLRERLATDISPIDVVPPIRLC 300  
 SFQRAVLDDLLAKT+KAL +YP K LVVAGGVANNQGLR+ELA +I+ I+VVIP IRLC  
 Sbjct: 241 SFQRAVLDDLLAKTKKALSRYPAKMLVAVGGVANNQGLRDLRLAQBIT-HIENVIPKRLC 299

25 Query: 302 GDNAGMICALAAATFEKFNFAKLKLNKPSLAPE 334  
 GDNAGMICALAAAT+K++FA++ LNKPSLAPE  
 Sbjct: 300 GDNAGMICALAAATYDKQHFANMBSLNKPSLAPE 333

25 SEQ ID 4926 (GBS371) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 7; MW 41kDa), in Figure 170 (lane 4 & 5; MW 55kDa) and in Figure 239 (lane 6; MW 55kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 7; MW 65kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1594

A DNA sequence (GBSx1688) was identified in *S. agalactiae* <SEQ ID 4929> which encodes the amino acid sequence <SEQ ID 4930>. Analysis of this protein sequence reveals the following:

Possible site: 33  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1027 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1595

A DNA sequence (GBSx1689) was identified in *S. agalactiae* <SEQ ID 4931> which encodes the amino acid sequence <SEQ ID 4932>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1307 (Affirmative) < succ>

-1782-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1596

- 10 A DNA sequence (GBSx1690) was identified in *S. agalactiae* <SEQ ID 4933> which encodes the amino acid sequence <SEQ ID 4934>. This protein is predicted to be L4171-60 protein. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have a cleavable N-term signal seq.

- 15 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 10135> which encodes amino acid sequence <SEQ ID 10136> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAC24656 GB:AB001274 L4171.5 [Leishmania major]  
 25 Identifies = 118/282 (41%), Positives = 167/282 (58%), Gaps = 4/282 (1%)  
 Query: 2 GGTQINQVVISMLASYEGVIAAETHVSSHEAGATIEFSGHKVLTLPSHNGKLLASEVAT 61  
 GGTQIN + S L +E VIA + GH+S+HE GAIE +GHKV+T P +GKL ++  
 Sbjct: 74 GGTQINLIACSLALRPWEAVIATQLGHISTHETGAIEATGHKVVTAPCPDGKLRVAD--- 130  
 30 Query: 62 YIETPYADGNQYHMFVPGMVIYISHPTEYGLYSKAESELSKICHHYQIPLFDGARLGY 121  
 IE+ + +HMF P +VYIS+ TE GT Y+K ELE++S CK + + LF+DGARL  
 Sbjct: 131 -IESALHENRSEHMFVPLKVIYISNITEVGTQYTKQLSDISASCKHGLYFLFDGARLAS 189  
 35 Query: 122 GLAAKDTVDVFPPTIAALSDFVYIGTGMGALAGEAVVFTKQNPQFTTIVKQRGALLAK 181  
 L++ D+ IA L+D+FYIG TK G + GER++ ++KQ GR+AK  
 Sbjct: 190 ALSSFVNDLTADTARLTDMFYIGATKAGMFGELIITINDALKPNARHLIKQRGALIMAK 249  
 Query: 182 GRLLGLAIFDRFPTDNLKYGKHAIDLABELKIIIEERGSFYLESPTNQOFTIVENTKL 241  
 G ILG+ F+ DNL+ ++G H+ +A LK LE G S +NQ F I+ENT +  
 40 Sbjct: 250 GWLLGIQFEVLMKDNLPFELGAHNSMGAAILKAGLEACGIRLAWPSASNLQFTILENTMI 309  
 Query: 242 ADLAKNVAYSFWKYYDHHFTVIRLATSWTSKREDVTALRNVL 283  
 A+L + E D ++RL TSW+T ++ VL  
 45 Sbjct: 310 AEANNDPEMYTVEPLKGTCTMRLCSTWATERKGRHRFVEVL 351

No corresponding DNA sequence was identified in *S. pyogenes*.

- SEQ ID 4934 (GBS648) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 8-10; MW 60kDa) and in Figure 186 (lane 6; MW 60kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 12; MW 35kDa), in Figure 140 (lane 10; MW 35kDa) and in Figure 178 (lane 7; MW 35kDa).

Purified GBS648-GST is shown in Figure 243, lane 6; purified GBS648-His is shown in Fig. 229, lane 7.



-1783-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1597**

A DNA sequence (GBSx1691) was identified in *S.agalactiae* <SEQ ID 4935> which encodes the amino acid sequence <SEQ ID 4936>. Analysis of this protein sequence reveals the following:

Possible site: 13  
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2279(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1598**

A DNA sequence (GBSx1692) was identified in *S.agalactiae* <SEQ ID 4937> which encodes the amino acid sequence <SEQ ID 4938>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3848(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12716 GB:Z99108 similar to ribosomal protein S14 [Bacillus subtilis]  
Identities = 67/89 (75%), Positives = 76/89 (85%)

```

Query: 1 MAKKS*IAKPKQKQLVQYAEIARRREIKKGGDYALRLKLPKDSNPNRLQNRDLIDGRPHA 60
MAKKS*AK K+Q+LVEQYA +RRELKKEGGDYAL KLP+DS P RL NR ++ GRP A
Sbjct: 1 MAKKSVAKRLKARQLVQYAGIRRELKKEGGDYALRLKLPDSAPGRLENRCMVTRPRA 60

```

```

Query: 61 YMRKFGMSRINRNFLAYKGGQIPGIKKASW 89
YMRKF MSRI FR LA+KGGIPG+KKASW
Sbjct: 61 YMRKFGMSRIAPRELAKGGQIPGVKKASW 89

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4939> which encodes the amino acid sequence <SEQ ID 4940>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-1784-

Identities = 73/89 (82%), Positives = 85/89 (95%)

Query: 1 MAKKSKIAKFKQKQKLVBQYAEELRRELKRGKGDYEAIRKLKPKDSNPNRLKNRDLIDGRPHA 60  
 MAKKSKIAX+QRQ +L+BQYA+LRR+LK KGDYB+LRKLK+DSNPNRLKNRDLIDGRPHA  
 5 Sbjet: 1 MAKKSKIAXQKQLCLTBQYADLRRLKAGKGDYBRLKRLKPKDSNPNRLKNRDLIDGRPHA 60

Query: 61 YMRKFGMSRINFRNLAYKGQIPGKIKASW 89  
 YMRKFG+SRINFR+LA+KGQ+PG+ KASW  
 10 Sbjet: 61 YMRKFGVSRINFRDLAHKQQLPGVTKASW 89

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1599**

A DNA sequence (GBSx1693) was identified in *S.galactiae* <SEQ ID 4941> which encodes the amino acid sequence <SEQ ID 4942>. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                         |                                        |
|-------------------------|----------------------------------------|
| bacterial cytoplasm --- | Certainty=0.5183 (Affirmative) < succ> |
| bacterial membrane ---  | Certainty=0.0000 (Not Clear) < succ>   |
| bacterial outside ---   | Certainty=0.0000 (Not Clear) < succ>   |

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA95931 GB:AL359989 galactose-1-phosphate uridylyltransferase  
 [Streptomyces coelicolor A3(2)]  
 Identities = 31/105 (29%), Positives = 51/105 (48%), Gaps = 4/105 (3%)

Query: 27 DKCPFC--DKSQLKILDVKDDMINVENKYPTL--EETQTLVIESMDHNGDISVYESK 82  
 D+CP C D +L +I D D++ EN++P+L + +V ++DH+ + SE +  
 30 Sbjet: 68 DQCPLCPDGERLSEIPDSAYDVVVFENRFPGLAGDSGRCEVVCFPSDHDASFADLSEK 127

Query: 83 MRQLLDYLLSKWQLMESGHYRSVVLRYNFGPLSGGSLRHPHSQI 127  
 R +LD + + V + N G G +L HPH QI  
 35 Sbjet: 128 ARLVLDATDRTSELHSLPSVQVFCFENRGAIGVTLGHFHQI 172

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1600**

A DNA sequence (GBSx1694) was identified in *S.galactiae* <SEQ ID 4943> which encodes the amino acid sequence <SEQ ID 4944>. Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

|                         |                                      |
|-------------------------|--------------------------------------|
| bacterial membrane ---  | Certainty=0.0000 (Not Clear) < succ> |
| bacterial outside ---   | Certainty=0.0000 (Not Clear) < succ> |
| bacterial cytoplasm --- | Certainty=0.0000 (Not Clear) < succ> |

A related GBS nucleic acid sequence <SEQ ID 10133> which encodes amino acid sequence <SEQ ID 10134> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1785-

>GP:BAB06998 GB:AP001518 unknown conserved protein [Bacillus halodurans]  
Identities = 186/410 (45%), Positives = 258/410 (62%), Gaps = 27/410 (6%)

5 Query: 4 YDTIIIGGSPAGMAAIISSNFYGNKTLTLEKNRRLGKKLAGTGGGRCNVTNNGNLDELLA 63  
++ I+IGGSPAG+MA++S+ +G + LL++K +LG+KLA +GGGRCNVTN LDEL+A  
Sbjct: 2 HEVTVIGGSPAGLAAVSAAEHKARVLLLDKGDKLGRKLAISGGGRCNVTNRMPLDELIA 61

10 Query: 64 GIPGNRGFLYSVFSQDFNHHDIINFPQDNGVTLKEEDHGRMFPTTDKSRITINALENKIKE 123  
IPGNRGFLYS FS F+N DII FF+ G+ LKED GRMF +DK+ T++ L +I +  
Sbjct: 62 HIRGNRGFLYSVFSVFNHEDIIRFFERLQALKEEDHGRMFVSDKATTVQTLARIND 121

15 Query: 124 LGGQIMTDTEVVSVK-KIGDSFYIKTKDTQFASDK-LIVTTGKSYSPSTGSGRGHDIAR 181  
LG + T+T V S++ G ++ K+ + K +IV TGG+S P TGSTG + A+  
Sbjct: 122 LGVTVKTNFAVASLEYDDGRIMVQLKNGERLKTIVTATGGQSVPTGTSGDAYFWAK 181

20 Query: 182 HFKLQVTDMEAAESPLLTDFP---HKKLQGISLDDVTLSE---EKHIITH---DLLFIYHF 232  
+T++ E P+ + P KKLQ+SL D+ LS K I TH D++PTHF  
Sbjct: 182 AAGHTITLTYTEVPITSAEPFIEQKQLGSLRDLIELSVYAFNGQIKTHDGMIFTHF 241

25 Query: 233 GLSGPAALRISSPVKGGSTIY-----LDVLPNWSVKEL-BIHFN---RREKSLGNA 280  
GLSGPAALR S +V Y +D+ P I + L + QN E +K+LK  
Sbjct: 242 GLSGPAALRCQYVVKALKKYKQPTIEMRIDLPTIPALALFQETIQNKASPKKALKTV 301

30 Query: 281 LKILLPERLAEPYAEEL--PEKVKQSVSKD--LEMLIQKLKLPILVTGKMSLAKSPVTK 336  
L+ + PER ++ E L + +SV+ + ++Q+LK V G +S+ K+FVT  
Sbjct: 302 LRGIAPEFLQYIYERLRIDSLKPCASVREHREIVQQLKSPSHVNGTLSISKAFVTG 361

35 Query: 337 GGVDLKBIKPTLESKQVAGLHFAGRVLDINAHTGGFNITSALCTGVAG 386  
GGV +KEI PKT+ SKK AGL F QVLDI+ +TGG+NIT A TG+ AG  
Sbjct: 362 GGVSVKRIEPTKMHSSKAGLFFQGVLDIHGTGGNITCAFTSGYTGA 411

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4945> which encodes the amino acid sequence <SEQ ID 4946>. Analysis of this protein sequence reveals the following:

35 Possible site: 23  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0448 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 308/386 (79%), Positives = 344/386 (88%)

45 Query: 1 MKHYDTIIIGGSPAGMAAIISSNFYGNKTLTLEKNRRLGKKLAGTGGGRCNVTNNGNLDE 60  
M YDTIIIGGSPAGMAAIISS++YG KTLLEKN+RLGKKLAGTGGGRCNVTN+GNLD  
Sbjct: 1 MTQYDTIIIGGSPAGMAAIISSSYGYKTLLEKNRRLGKKLAGTGGGRCNVTN+GNLDV 60

50 Query: 61 LLAGIPGNRGFLYSVFSQDFNHHDIINFPQDNGVTLKEEDHGRMFPTTDKSRITINALENK 120  
L+LAGIPGNRGFLYSVFSQDFNHHDI FF+NGV LKEDHGRMFPTTDKSRIT+ALE K  
Sbjct: 61 LLAGIPGNRGFLYSVFSQDFNHHDI IAPFENGVLKEDHGRMFPTTDKSRITIDALEKK 120

55 Query: 121 IKELGQIMTDTEVVSVKKIGDSFYIKTKDTQFASDKLIVTTGKSYSPSTGSGRGHDIAR 180  
IK LGQ++T TEVVSVKK D FY+K+ D F KLIVTTGKSYSPSTGSGRGHDIAR  
Sbjct: 121 IKALGQVLTSTEVVSVKKQDDLPLYLSADQITFQKLVITVTGKSYSPSTGSGRGHDIAR 180

60 Query: 181 RHFKLQVTDMEAAESPLLTDFPHKKLQGISLDDVTLSEPHKIIITHDLFTHPGLSGPAAL 240  
RHFKL VTD+EAESPLLTDFPHK LQGISLDDVTL++KH+I+THDLFTHPGLSGPAAL  
Sbjct: 181 RHFKLVTDLAASPLLTDFPHKVLQGISLDDVTLSEYKHIVITHDLLFTHPGLSGPAAL 240

65 Query: 241 RISSPVKGGSTIYLDVLPNWSVKELIHFNQREKREKSLKALLPERLAEPYAEELPEK 300  
R+SSPVKGGG LD LP++S +L + ++K++KNAK ILPER+A+F +ED PEK  
Sbjct: 241 RLSSPVKGGSTIARLDPLFLSLTDLTAYLSDQRDNKIKNAKGLPERVADFLSEDPYK 300

Query: 301 VKQSVKIDLEMLIQKLKLPILVTGKMSLAKSPVTKQGVLDKRIKPTLESKQVAGLHFA 360

-1786-

VKQ+S K + L+ K L K L I +TGRMSIAKSFVTKGGVDLKEINPKTLASKKV GL+FA  
 Sbjct: 301 VKQLSPKQEKELLDLKLHLQIPITGRMSIAKSFVTKGGVDLKEINPKTLASKKVPLGLYFA 360

Query: 361 GEVLDINAHTGGFNITSALCTGWWAG 386  
 GEVLDINAHTGGFNITSALC+GW+AG  
 Sbjct: 361 GEVLDINAHTGGFNITSALCSGWIAG 386

SEQ ID 4944 (GBS196) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 3; MW 44.5kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 4; MW 69.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1601

A DNA sequence (GBSx1695) was identified in *S. agalactiae* <SEQ ID 4947> which encodes the amino acid sequence <SEQ ID 4948>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1550 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10131> which encodes amino acid sequence <SEQ ID 10132> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA73267 GB:Y12736 orfX [Lactococcus lactis subsp. cremoris]  
 Identities = 51/173 (29%), Positives = 87/173 (49%), Gaps = 20/173 (11%)

Query: 19 KTVSELAEILGVSRQAMNNRV-KTLPEECVEK---NSKGVTVVNRDGLKLEIYKKTIL 74  
 KT+ ELA+ LGVS+Q + N++ K E+ V+ V+N G + KKT+  
 Sbjct: 6 RTIKELADELGVSKQTIRNKIDCFREKVFQVTIKINGNTLVINNAGY----SLKKTILQ 61

Query: 75 EREPIDEEASRRLEILVDERNTEITRLYEQLAKDIQIASKDQLHVKDIQIAEKDRQ 134  
 + + + + + I L EQL K+ Q++ KD+QL KD QI++  
 Sbjct: 62 NDTAQTAKTLQNDTAQTKL-----ICFLEQLDKKEQQLSVKDKLENDTQISQML 115

Query: 135 LDQQQQLTLTAMSDTORLQLELNKA-----EVEIEQAEKEEIKQIELEAVK 181  
 LDQQQ+L L + + + E+NR KA +++++ + E +E+E +K  
 Sbjct: 116 LDQQQLALQDKKILERYKSEINELKALMPREDMKDGSSIRGEAQREIRFLK 168

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4949> which encodes the amino acid sequence <SEQ ID 4950>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3951 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 132/194 (68%), Positives = 154/194 (79%), Gaps = 4/194 (2%)

-1787-

Query: 1 MIFYYKKI---STKKEVMTVEKTVSELAELIGVSRQAMNNRVKTLPEECVKNKSGVTVV 57  
 M+ F +I S KEE M +EKTVSELA+ILQVSRQA+NNRVK+LPKE ++KN KGVTVV  
 Sbjct: 1 MVLFLIRI PDSDDKEENNGI EKTVSELAELIGVSRQAVNNRVKSLPDELDHNEKGVTVV 60

5 Query: 58 NRDLGILKEEYIKTILKEEPIDEEASRELETLVDEKNTETIRLYEQLKAKDIQASK 117  
 R GL+KLEEYIKTII +EPI EE +RELETLVDEKNTETIRLYEQLKAKD Q+ASK  
 Sbjct: 61 KRSLGVLKEEYIKTIFDEPISSEETKQRELETLVDEKNTETIRLYEQLKAKDAQASK 120

10 Query: 118 DEQLHVVD+QIARCKQLDQQQQITLTAMSDTQRLQLEINAKAEVVEIQAKKEKIQEL 177  
 DEQ+ VKD+QIARCKQLDQQQQIT AM D +L+LEL EAKAE + + + E++Q  
 Sbjct: 121 DEQMRVVDQIARCKQLDQQQQITAKAMADKRTLKLELEAKAEANQAR-IQVEVQAR 179

Query: 178 EAVKSGFFGRFPNK 191  
 KK FF R F K  
 15 Sbjct: 180 VGGKSGFFTRLFAK 193

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1602

- 20 A DNA sequence (GBSx1697) was identified in *S.agalactiae* <SEQ ID 4951> which encodes the amino acid sequence <SEQ ID 4952>. Analysis of this protein sequence reveals the following:

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.2157 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06137 GB:AP001515 DNA polymerase III (alpha subunit)  
 [Bacillus halodurans]  
 Identities = 31/87 (35%), Positives = 52/87 (59%), Gaps = 1/87 (1%)

- 35 Query: 13 EYIAFDLEFNTVGE-RSHIIQVSAVKYNNHQIALFDITYVHTKVPQLSPINGLTGITARD 71  
 E++ FD+E + ++ II++++AVK N + I F+ + PL + I LTGIT  
 Sbjct: 418 EFVVFVDVETGLSAVYNNKIELAAVKVINGEIIIDRFERFADPPEPLNTIIEILTGITDDM 477
- 40 Query: 72 IIGAPKIEIVLIDFQSFVGDPLIGYN 98  
 + G P++E VL +F +F+GD L+ +N  
 Sbjct: 478 LKGQPEVEQVLNEPHAFIGDAVLVAHN 504

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4953> which encodes the amino acid sequence <SEQ ID 4954>. Analysis of this protein sequence reveals the following:

- 45 Possible site: 29  
 >>> Seems to have no N-terminal signal sequence
- 50 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3427 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 136/200 (68%), Positives = 159/200 (79%)
- 55 Query: 3 FLGSEIMKQLQEIYIAFDLEFNTVGE-RSHIIQVSAVKYNNHQIALFDITYVHTKVPQLSPIN 62  
 FL E MK L YIAFDLEFNTV + SHIIQVSAVKY +H+E+ FDTYV+T VPLQSPIN  
 Sbjct: 9 FLEENMKHLDYIAFDLEFNTVNVSHIIQVSAVKYDHHKEVDSPDTYVYTVVPLQSPIN 68

-1788-

Query: 63 GLTGITARDIIGAPKIEIVITDFQSFVGDITLIGYNGYKSDLLPVLVMSLDLTSCQVNDL 122  
 GLTGIT\* 1 PK+E V\* F\*+FVG+ FLIGYN KSDL\*+L ENGLDL QYQ+DL  
 Sbjct: 69 GLTGTSDKTAASPKVEVMRAFINFVGHLPFLIGYNAQKSDLPILAENGILDRDQYQIDL 126

Query: 123 YDRAPVPRSTDLNGIVNKLITTVADPLGIGKKAHNSLEDARMTARVYRKFLDLDENKIYL 182  
 +DER+ RRS DLNGI NL+L TVA FLGIGK+ HNSLEDARMTA +Y+ FL+ D NK YL  
 Sbjct: 129 FDEAYDRSADLNGIANLRLQVATVFLGIGKRGHNSLEDARMTAVIYKSFLETITNKAYL 188

Query: 183 KQKKEVAVDSFFATLGNLFD 202  
 QQ+EV D+FFA LG+ FD  
 Sbjct: 189 SQCEVTTIRFPAALGDFD 208

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 15 Example 1603

A DNA sequence (GBSx1698) was identified in *S. agalactiae* <SEQ ID 4955> which encodes the amino acid sequence <SEQ ID 4956>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence

20 INTEGRAL Likelihood = -12.10 Transmembrane 143 - 159 ( 136 - 166)  
 INTEGRAL Likelihood = -4.73 Transmembrane 169 - 185 ( 168 - 188)

----- Final Results -----

25 bacterial membrane --- Certainty=0.5840 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB42766 GB: AL049841 transcriptional regulator [Streptomyces  
 coelicolor A3 (2)]

30 Identities = 46/141 (32%), Positives = 71/141 (49%), Gaps = 11/141 (7%)

Query: 5 YSTGDLAKGAGVTVRTVYQYDKRGILSPSELSEGGRRVYSIADLEKLQIILYLEDLDFSI 64  
 YS G +A AGVTVRT+ +YD G+L PSE S G R YS ADL++L+QI++ R+L F +  
 35 Sbjct: 3 YSVGQVAGFAGVTVRTLHYDDIGLLVPSERSAGHRRYSADLDRQLQILFYRELGFPL 62

Query: 65 DNINKLFTEDNASQILELFLQVQIRELR-----AIDSKOKLKDKNVLLKTVKQD 116  
 D + L + A L Q ++ R+ R++ + +NL ++  
 40 Sbjct: 63 DEVAALLDDFAADPRAHLRRQHELLSARIGLQKNPAAVEQAMEARSNGINL--TPBEK 119

Query: 117 SKTLGVLSDIVMERKRWG 137  
 + G EE + +WG  
 Sbjct: 120 FEVFGDFDPPQYEEVRRWG 140

45 There is also homology to SEQ ID 1712.

SEQ ID 4956 (GBS372) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 8; MW 55kDa).

GBS372-GST was purified as shown in Figure 215, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 50 vaccines or diagnostics.

**Example 1604**

A DNA sequence (GBSx1699) was identified in *S.agalactiae* <SEQ ID 4957> which encodes the amino acid sequence <SEQ ID 4958>. This protein is predicted to be cyclopropane-fatty-acyl-phospholipid synthase (mma2). Analysis of this protein sequence reveals the following:

```

5      Possible site: 53
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3145 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:AAD07482 GB:AB000557 cyclopropane fatty acid synthase (cfa)
      [Helicobacter pylori 26695]
      Identities = 167/397 (42%), Positives = 254/397 (63%), Gaps = 14/397 (3%)

      Query: 2   VMDSLIIKQLIKSTFDIPLQVTYPNGNIETYNNGSNPHVLLKLNKFSVSELSKDPISVLG 61
      ++ ++K + K + QV + + ++ +P LK+++ S++ KD S+ +
20      Sbjct: 1   MISKFLIKSMFKQMKNGDYQVVFWDNSVYRNGEHSKFTLKIHRPLKFSDDIKDMSLTIA 60

      Query: 62   EAVMDGDIETIYSGIQELILSAY-ROGDSFLRNSKFSKLIPKQPHDKHKSSEDIQKHYDIG 120
      EA MDG I+I GS+ E++ S Y + L +K I K + S+I KHYD+G
25      Sbjct: 61   EAYMDGVIDIRGSMDEVMSLYLQ/TNYRHLHCHDHNAIAIQKPIKES----SNISKHYDLG 116

      Query: 121  NDFYKMLDDTWTYS CAYFKHENDSLEQAQLNKHVHILNKLNAQGGKLLDIGCGWTLI 180
      NDFY +WLD+T++YS CAYFK ++D+L AQL K+ H L KL+ +PG KLLDIGCGWG L
30      Sbjct: 117 NDFYSIWLDDTSLSYSCAYFKKDDDTLHAAQLKLDHTLKKLHLKPEKLLDIGCGWYIS 176

      Query: 181  ITAAKEYGLNATGILTSEQASFTIKRIKESGLENNKVTILKDYRDI---RETYDYITSV 237
      + AA+YEG GIT+S EQ KR++E GLR+KVT++ +Y+D+ +D +S +
35      Sbjct: 177 VKAAQYGAENVGITISSEQYKQANKRVQELGLEDKVTIKLHNYQLDGRLYRFDKVVSV 236

      Query: 238  GMFERHVGKENSQYFQTISKRLNGLALHIGITQGVGNHSGSGTNSWINKYIPFGGYIP 297
      GMFERHVGK+NL YF+ + + L G+ L+H I G TN+W++KYIPFGGY+P
40      Sbjct: 237 GMFERHVGKDNLPFYKKVKEVLKRGGMELLSILCCFEGK---TNAWDKYIPFGGYLP 292

      Query: 298  RLITENLNHIASGLQIADLEPLRRHYQKTLLELWTKNHNALPEVQK-THDKRFINMDWLY 356
      L E ++ ++ + E LR HY KTL+W NF++ + L +V++ ++D+RFI MDWLY
45      Sbjct: 293 SLREVMVSMSECDPHLLMAESLRHYAKTLDTWRNHNHNLQVRLSYDERFIRMWLY 352

      Query: 357  LQSCAASFESGNIDIPQYLLSKGVSKDTPWMPTRDYM 393
      L++CA++F G+ D+PQ LL+ V +T P+T++Y+Y
      Sbjct: 353 LRTCAAPKRVGSADLPOLL/NSVD-NTFPLTKETIY 388

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1605**

50 A DNA sequence (GBSx1700) was identified in *S.agalactiae* <SEQ ID 4959> which encodes the amino acid sequence <SEQ ID 4960>. Analysis of this protein sequence reveals the following:

```

      Possible site: 35
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
55      bacterial cytoplasm --- Certainty=0.4903 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-1790-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11796 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 44/97 (45%), Positives = 60/97 (61%)

```

Query: 1  MNNMNNMRRQAKLQKQMEQKQADLAASQFTGKSAQELVTVFTGDKKLISIDYKRAVVD 60
      M NMQ MM+Q QK+QK M + Q +LA      G + +VTV  G K+++ + KE VVD
Sbjct: 5  MNNMNNMRRQAKLQKQMEQKQADLAASQFTGKSAQELVTVFTGDKKLISIDYKRAVVD 64

Query: 61  FEDIETLQDMTTQAINDALSOVDDATKKIMGAFAGKM 97
      PEDI+ LQD+  A N+AL +VD+ T + MG F M
Sbjct: 65  FEDIMLQDLVLAATNEALKKVDIEITNETMQFTKGM 101

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4961> which encodes the amino acid sequence <SEQ ID 4962>. Analysis of this protein sequence reveals the following:

Possible site: 35  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.4451 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/99 (84%), Positives = 94/99 (94%)

```

Query: 1  MNNMNNMRRQAKLQKQMEQKQADLAASQFTGKSAQELVTVFTGDKKLISIDYKRAVVD 60
      MNNMNNMRRQAKLQKQMEQKQADLAASQFTGKSAQELVTVFTGDKKLISIDYKRAVVD
Sbjct: 1  MNNMNNMRRQAKLQKQMEQKQADLAASQFTGKSAQELVTVFTGDKKLISIDYKRAVVD 60

Query: 61  FEDIETLQDMTTQAINDALSOVDDATKKIMGAFAGKM 99
      FED+ETLQDMTTQAINDAL+Q+D+ TTK +GAFAGK+PF
Sbjct: 61  FEDVETLQDMTTQAINDALVQIDETTTKTLGAFAGKLPF 99

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1606

A DNA sequence (GBSx1701) was identified in *S.agalactiae* <SEQ ID 4963> which encodes the amino acid sequence <SEQ ID 4964>. Analysis of this protein sequence reveals the following:

Possible site: 17  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.3963 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



-1791-

**Example 1607**

A DNA sequence (GBSx1702) was identified in *S.agalactiae* <SEQ ID 4965> which encodes the amino acid sequence <SEQ ID 4966>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 48
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL    Likelihood = -2.76    Transmembrane    21 - 37 ( 19 - 39)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.2105 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10129> which encodes amino acid sequence <SEQ ID 10130> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1608**

20 A DNA sequence (GBSx1703) was identified in *S.agalactiae* <SEQ ID 4967> which encodes the amino acid sequence <SEQ ID 4968>. Analysis of this protein sequence reveals the following:

```

      Possible site: 36
      >>> Seems to have no N-terminal signal sequence

25      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1783 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1609**

35 A DNA sequence (GBSx1704) was identified in *S.agalactiae* <SEQ ID 4969> which encodes the amino acid sequence <SEQ ID 4970>. This protein is predicted to be probable 1,4-dihydroxy-2-naphthoate octaprenyltransferase. Analysis of this protein sequence reveals the following:

```

      Possible site: 32
      >>> Seems to have no N-terminal signal sequence

40      INTEGRAL    Likelihood = -8.76    Transmembrane    239 - 255 ( 219 - 260)
      INTEGRAL    Likelihood = -8.33    Transmembrane    221 - 237 ( 219 - 238)
      INTEGRAL    Likelihood = -6.74    Transmembrane    91 - 107 ( 89 - 113)
      INTEGRAL    Likelihood = -6.32    Transmembrane    39 - 55 ( 35 - 59)
      INTEGRAL    Likelihood = -3.77    Transmembrane    111 - 127 ( 111 - 132)
45      INTEGRAL    Likelihood = -2.97    Transmembrane    144 - 160 ( 143 - 161)
      INTEGRAL    Likelihood = -1.28    Transmembrane    275 - 291 ( 275 - 291)
      INTEGRAL    Likelihood = -0.59    Transmembrane    177 - 193 ( 177 - 193)

```

-1792-

## ----- Final Results -----

bacterial membrane --- Certainty=0.4503 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15875 GB:Z99123 alternate gene name: ipa-6d-similar to  
 quinone biosynthesis [Bacillus subtilis]  
 Identities = 75/290 (25%), Positives = 139/290 (47%), Gaps = 15/290 (5%)

Query: 5 IFLSELVEMKAKTASVLFFLIGLCFSAYYNSVHPVYVGLFVAMFLNNMFVDIWNVNDY 64  
 I +L TAS +P L+G + +Y +++ + P +++ + +++N Y D+  
 Sbjet: 21 ILAQQLTRPHLTASFVPVLLGLTVMAMPYVKVLDLLFLAMLPSCLMI-QIATNLNPEYIDF 79

Query: 65 RNAVDL-DYKNDIINIGRENLSLRQIEVIMASLVTSSMIGLVNLSQVGLPLMLMGLPCF 123  
 + +D + I R + + I + + + +G+ + + L +GL  
 Sbjet: 80 KRGLDTABSVGIGGAIVRHGMKPKTLLQALASYGIALLLGVYICASSSWWALALIGLVGN 139

Query: 124 GIGVLYSPGPRPLSSPLGEVFSGLTMGMISLICVILTYQNFSDIILNLSKIFLISLP 183  
 IG LY+ GP P++ P GE+PSG+ MG + LI ++ T D +N+ I LIS+P  
 Sbjet: 140 AIGVLYTGGLPIATYTPFGELPSGICMGSVFVLSFFIQT-----DKIMQSG-LIISIP 192

Query: 184 NTLNLANLMLANNLCDEEDENHRYTLVHYTCIRGOLLPAISNSIALLAIVPEFLPL 243  
 + + + L+NN+ D EED+K R TL G +G + L A S ++A + +V + G  
 Sbjet: 193 IAILWGAINLSNNIRDIIEEDKKGKRTLAAILAGHIGAVILAASEFAVAYVWVGLVITGA 252

Query: 244 APVTVLSLLLIPIFYKTKLLKQKQVKRETFVCAVRILALGSATQVLT 293  
 A + + L +P + K Q ++ I++ S Q T+  
 Sbjet: 253 ASMLPVVFLSVKPKVQAVIGFVQGNEMPM-----MIVAMKSTACTNTF 296

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 1610

A DNA sequence (GBSx1705) was identified in *S.agalactiae* <SEQ ID 4971> which encodes the amino acid sequence <SEQ ID 4972>. Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.22 Transmembrane 155 - 171 ( 154 - 171)

----- Final Results -----

bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15200 GB:Z99120 similar to NADH dehydrogenase [Bacillus subtilis]  
 Identities = 178/403 (44%), Positives = 249/403 (61%), Gaps = 7/403 (1%)

Query: 3 EILVLGAGYAGLKAVRNLOKQSG--DPHTLVDMNDYHYEATLGEVAAGSQPKKITFP 60  
 I++LGAGY GL V L K G D ITLV+ ++HYE T +HE +AG+ ++ +  
 Sbjet: 7 KIVILGAGYGGIMTVTRITKYVGENDADITLVNKENYHYETTWHEASAGTILHIDRCRYQ 66

Query: 61 IKDVININKVNFMDQEVLRDAENKTVIVKNNGLHYDYYVVALGVSEVETFGIKGAMNA 120  
 IKDVIN ++VNF+QD V + + K V + N GRV XDY+V+ LG V ETFGIK G E A  
 Sbjet: 67 IKDVINQSRVNFVQDTVKATKIDEKKVVLAN-GEQDYLVTIGLAVPETFGIKGLKLYA 125

Query: 121 LQMTNISQAEINHHNHHVIMKLYRETCIE--NLKLLVQAGTFTGIELAGAMVDERPPIYA 178  
 + NI+ + + HI Y ++ + L +V GAGFTGIE G + P+

-1793-

Sbjct: 126 PPTANINTSRLLRSHIELQFATYNTAEKRPRDRIITVGGAGFTGIEPLGELAAVPELC 185

Query: 179 ALAGVKEPQIEIICVAATRILPMFDELAQYGVNLKIDGLINMLGSMIKIRKGVVY 238  
V + IICVBA +LP FD SL Y V+ +++ G+ +G+ ++E P V

5 Sbjct: 186 KEYDVRSLVRILICVAAPVLGDFDELVDYAVHLEENGVEFKIGTAVQBCCTPESVRV 245

Query: 239 GTSKEDEELKSITAGTIITWTGVSQFVMSGSGFDQRRGRVMVNSLDLPKYNVIVIGD 298  
G K+DEE + I + T++W GV G P++ E+GF+ RGRV VN DLR P +DNV+++GD

10 Sbjct: 246 G--KCEBEPQIKSQTVVAAQVRGHPIVEAGFENRGRVKNVPLARAGHDNVFLIGD 303

Query: 259 VSAFMOTESGRFFFTTAQIATRMGAHVAKNLLAQIKGRATEDFSYSPQQTVASVGNTHGL 358  
S FM+ ++ RP+P TAQIA + G VAKNL IKG E+P +GTVAS+G ++

15 Sbjct: 304 SSLPMNEDTERPYPTTAQIAMQQITVAKNLGRLINGGLEEFKPDIGTVAISLGHNAV 363

Query: 359 GVGKTKIKKYPASVNMKKIIMNKSIVDMGGLKELLANGRPDLY 401  
GVV K+K PAS MKK+I N+SL +GGL L KG+P +

Sbjct: 364 GVVYGRKLKGTASPMKKVIDNRSLFMIGGLGUILKKKGKPKFF 406

There is also homology to SEQ ID 4666.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1611

- A DNA sequence (GBSx1706) was identified in *S. agalactiae* <SEQ ID 4973> which encodes the amino acid sequence <SEQ ID 4974>. This protein is predicted to be cytochrome d ubiquinol oxidase, subunit I (cydA-1). Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -6.64 Transmembrane 19 - 35 ( 15 - 38)  
INTEGRAL Likelihood = -5.73 Transmembrane 226 - 242 ( 222 - 244)  
INTEGRAL Likelihood = -4.94 Transmembrane 130 - 146 ( 126 - 149)  
INTEGRAL Likelihood = -4.83 Transmembrane 429 - 445 ( 422 - 446)  
INTEGRAL Likelihood = -3.77 Transmembrane 55 - 71 ( 53 - 74)  
INTEGRAL Likelihood = -3.56 Transmembrane 342 - 358 ( 340 - 359)  
INTEGRAL Likelihood = -1.06 Transmembrane 89 - 105 ( 89 - 106)  
35 INTEGRAL Likelihood = -0.59 Transmembrane 186 - 202 ( 186 - 202)

----- Final Results -----

bacterial membrane --- Certainty=0.3654(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15902 GB:299123 cytochrome bd ubiquinol oxidase (subunit I)  
[Bacillus subtilis]

45 Identities = 246/470 (52%), Positives = 319/470 (67%), Gaps = 12/470 (2%)

Query: 6 IARQFAMTYVPHFFVPVPTTIGTCLVALMETMYVITONEEYKLLTFWGNMILLSPVAV 65  
LAR QFA TT+PHF FVP +IG +VA+MET+Y++ KNE Y K+ KFWG++ L++FAVG  
Sbjct: 6 LARIQFASITLPHFLFVMSIGLVFMVAKMETLYLVKKNELLYKMAKFWGHGLINFAVG 65

50 Query: 66 VVNGIIGQFQFGMWSGDSYRFVCDIFGAFALAEALLAFMBSYTFGLNMPTWDNKISKK 125  
VVNGIIGQFQFG+MWSGDSYRFVCD+FGAFALAEALLAFMBS F+GLN+P ND ++ KK  
Sbjct: 66 VVNGIIGQFQGLNWSGDSYRFVGVDFGAFALAEALLAFMBSIIFGLNFMGD--RLFKK 123

55 Query: 126 LHVTFIMLVFVSGLSASAMWLLTANSFMQHPYGVSVVGRQAQMDPLAVKNPQFFVEPTH 185  
+H +HVN PG++MS+ WILLTANSFMQ FVG+ +NGRA+M DP AL+ NEQ + EF H  
Sbjct: 124 IHALCIWLVSPGTIMSSFWILLTANSFMQHPYGVGTIKNGRAMMDGALITNPQLAVVEPH 183

60 Query: 186 VIRGALTMGGTVVAGMSAFRLIKSEQLKDTTVELYKYSVRIGLIVALLGSIWWSGDLQ 245  
VIRGA+ G +AG+SAF+LEK ++ V +K+S ++ ++V L + V G +Q

-1794-

5  
 10  
 15

```

  Sbjct: 184 VIFGALATGAFFIAGVSAFKLLKKE-----VPPFKQSKFLAMIVGLCAGLGVGLSGHM 238
  Query: 246 MKALIHDPMKFAAMRGSDYEDGQDPAAGVAVAMANEASHKQVFGIKI PYMLSILSYGKPS 305
      + L+ QPMK AA BG +EDGQDPAW+ A + K IK+PY LS L+Y K S
  Sbjct: 239 AEHLMEQPMQKMAASBSGLWEDGQDPAWITATITDTNKEKSSHEI KPVYALSYLAQKFS 298
  Query: 306 GSVKGMGTANKELVAKYKGDNYYPWNLLYFGPRITMAAMGTAIMGVSVLGLFTRKKKPI 365
      GSVKGM T E YGK +Y P V F+ FR M G ++ ++ GL+L R+KK
  Sbjct: 299 GSVKGMKTLQAEYRKYKGDYIPFVKITFWSPRIMVGVVMIILALGLGLMIRK-- 356
  Query: 366 LYKHKWMLNIVALITTAFFLANITGMITVTEQGRYPWTVYGLFKRIKDSVSPNVSVASLFS 425
      L KW L I+ PFLAN+ GWI+TE GR PWTV GL SVSPNV+ SL S
  Sbjct: 357 LENSKWYLRIMIALISLFFFLANSAGNIMTEIGRQFWTVMGLMITAGSVSPNVITAGSLFS 416
  Query: 426 NTVYFLLEGLLAVNMLSLITRELKKGPEYEDHGHGAYTSDIPFEGAY 475
      + +++ L ++L IRE+KKG E+++ HH S DPF + Y
  Sbjct: 417 IIAFGVMYMLGALLVLEFIREIKKGAHDN---HHVVPVSTDPPSDEVY 463
  
```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1612

A DNA sequence (GBSx1707) was identified in *S.agalactiae* <SEQ ID 4975> which encodes the amino acid sequence <SEQ ID 4976>. This protein is predicted to be cytochrome oxidase subunit II (cydB-1).

- 25 Analysis of this protein sequence reveals the following:

30  
 35

```

  Possible site: 22
  >>> Seems to have an uncleavable N-term signal seq
  INTEGRAL Likelihood = -14.49 Transmembrane 226 - 242 ( 220 - 250)
  INTEGRAL Likelihood = -8.12 Transmembrane 254 - 270 ( 250 - 282)
  INTEGRAL Likelihood = -7.64 Transmembrane 198 - 214 ( 196 - 218)
  INTEGRAL Likelihood = -6.95 Transmembrane 85 - 101 ( 76 - 103)
  INTEGRAL Likelihood = -6.74 Transmembrane 6 - 22 ( 1 - 27)
  INTEGRAL Likelihood = -6.16 Transmembrane 300 - 316 ( 298 - 322)
  INTEGRAL Likelihood = -5.36 Transmembrane 119 - 135 ( 117 - 143)
  INTEGRAL Likelihood = -4.04 Transmembrane 159 - 175 ( 155 - 178)

  ----- Final Results -----
  bacterial membrane --- Certainty=0.6795(Affirmative) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the GENPEPT database.

45  
 50  
 55  
 60

```

  >GP:CB15901 GB:Z99123 cytochrome bd ubiquinol oxidase (subunit II)
  [Bacillus subtilis]
  Identities = 158/331 (47%), Positives = 223/331 (66%), Gaps = 1/331 (0%)
  Query: 1 MSALQFFWFLLGLLPSGFFLGGDFGVGMVAVTLTHANEKQGVETIGPVWDNEVW 60
      M++L WF L+ +LE GFFELGGDFGVGMA + L HNE Z+ ++ TIGE WD NEVW
  Sbjct: 1 MASLHDLWFLVAVLFGVFFLGGDFGVGMATRFGLHNELESRVLIINTIGFWDANEVW 60
  Query: 61 LITGGGAMPSPFYWASLPSGYILLITLLITFLGLIRGVSPFPRHVPARK-KQFVAVTL 119
      LLTG GA+FA+FP WYA++ SGY++ + +L L+ RGV+SEFR KV K + W+W +
  Sbjct: 61 LITGAGATFAAFDNNWATMLSGYIIPFVILLALMGKGVAFPRGKVDHLKWKVKVNEVW 120
  Query: 120 TIGSAIVPFFFGIMFSLIQKPLDASGSLGQFSDYFPIPSLVGVGVAMVLLAYLHGLAY 179
      GS I PF G++F +L +GMT+DA N+ A SDY N++S+QGV + L+ + HSL +
  Sbjct: 121 FFGSLIPPFVGLVLTLLTFRGMPIDADMNIHAEVDYINVYSILQGVTVTLCCQKSLAF 180
  Query: 180 IALKTGPIRERARNYAQLLYWVLYLGLALPAVLVLYFKTDFPSNHPITVTIMVLIIVLA 239
      I L+T G++ RAR AQ + V+++ + FA L ++TD F+ +T + ++IV+
  
```

-1795-

Subjct: 181 ITLRITGDLQRARKMAQKINGVVFVAFLAFAALSAYQTMFTRGEITIPAVLIVICF 240

Query: 240 VLAHASTFKGARMTAFLASGLSLVSVVLLFQGLFPRVMISSIPKYDLLQNASSTPYT 299  
+LA K + F +G L V ++F LFPFVM+SS+ YDL + NASS Y+

Subjct: 241 MIAAVFTRKKKGWTFGMTIGALALTVMIFITSLFPRVMVSSLHAYDLTVANASSGDYS 300

Query: 300 LKVMSTVAITLVPFVLAAYTAWAYYIFRKRIT 330

LKVMGI A+TL+PFV+ W+YY+FRKR++

Subjct: 301 LKVMSTAAITLLFPVIGSQTWSTYVFRKRVS 331

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1613

- 15 A DNA sequence (GBSx1708) was identified in *S.agalactiae* <SEQ ID 4977> which encodes the amino acid sequence <SEQ ID 4978>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

20 ----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1614

- 30 A DNA sequence (GBSx1709) was identified in *S.agalactiae* <SEQ ID 4979> which encodes the amino acid sequence <SEQ ID 4980>. This protein is predicted to be transport ATP-binding protein cydc (cydD). Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -16.82 Transmembrane 158 - 174 ( 144 - 182)

INTEGRAL Likelihood = -6.48 Transmembrane 15 - 31 ( 14 - 34)

INTEGRAL Likelihood = -5.31 Transmembrane 243 - 259 ( 238 - 266)

INTEGRAL Likelihood = -2.55 Transmembrane 136 - 152 ( 134 - 152)

INTEGRAL Likelihood = -0.48 Transmembrane 263 - 279 ( 263 - 279)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.7729 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 45 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15900 GB:299123 ABC membrane transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 279/569 (49%), Positives = 401/569 (70%), Gaps = 6/569 (1%)

50 Query: 2 LDKAVMLSGTHKLLGLAGLDVLAQIFIGQAYYLSLSITGLWESQKLSSTQVYILFM 61  
+ K + R G+ ++L L+ L ++Q II QA +LS ++TGL+ G+ ++S I F+

Subjct: 1 MSKDLFRYKMGKRILTLITCLTLIQTAITMQAEWLSSEAVTGLPNKGKGTSLLEPVIGFPL 60

-1796-

Query: 62 VSYLGRHVVIDYIKNRKLDQFSTAQSSLLRRQLDOKLFDLGPVKVQEGSTGVMVYMALDGV 121  
 ++++ RH + + + + + LR+ LD+LF LGP+ +++GTG +VTA++G+  
 Sbjet: 61 IAFIARHGHTVARQIKYIVQAARTGADLRKSFLOQLFRLGPRFAKKSGTGQMVYIAMEGI 120

Query: 122 SILVENYLRVLNKMIMMSIIPWILYIFLFDIESGAILLVFLLIIFMILGVAAQAK 181  
 S YL L L K M+M+I+P ++ Y+P+ D S IL+ P++IIPMI+LG AQ K  
 Sbjet: 121 SQFREYLELFLPKVMGMAIVPAVAVVITYVFFQURTSAILVRAAMPILLIIFMILGLVAVQRK 180

Query: 182 ADKQTESYQVLNSHIFDLSRGIDTLKYPLQSKFYKNGSIYQTSSEPRKATMTGLIGLST 241  
 AD+Q+SYQ LSNIF+DSLRG++TL++ GLSK + K+I+ SE +RKATMTGL++ L6+  
 Sbjet: 181 ADKQTESYQRLSNHIFVDSLRLGLETLRLFLGLSKHSHKNIPTVSEIRKATMTSTRAVFL6S 240

Query: 242 FALDFFFTLSIAIVAVFLGLRLNMQYILLPALTLILISPEYFLPVRDPSDDVHATLDGK 301  
 FALDFFT L6H+ VAVFLGLRL++ I L FALT LIL+PEYFLPVR+ ADVHATL+G+  
 Sbjet: 241 FALDFFFTLMSVAVAVFLGLRLDGLLGLPALTLILILAFSEYFLPVRVGVNDVHATLNOQ 300

Query: 302 NAFQAQKVLNKGIGKE-QLVIDDWSKESRLDENIAIAYDQKRVVEDVTIRFSGQKV 360  
 A + IQ++L+ G K E L ++ WS + L L +++ R V D+ L F+G +K+  
 Sbjet: 301 EAGKTIQELSGQFKEETPLQLEAWSQDELKLGVSVG----RSVSDIHL6FKGJKKI 356

Query: 361 ALWVGSGSGKSLINLISGLFLGPNWGLKVDGREVINLQDMDHMQMITYIPQVFPFMS 420  
 ++G SG+GK+LI++L GFL PD G ++V+G ++L W K ++XIPQ PY+P+ +  
 Sbjet: 357 GIIGASGAGKSLIDILGGLFEPDGGMIENVNTRSHLQDQSGMQLNLYI6QHFIYFDOT 416

Query: 421 LRDNITFYPTASDEEVVRAIHMGVGLDLSLSELPGDLETIGNARPLSGGQAQIALAR 480  
 L+NI FY P+AS E+ RA GL L++ LPDGLR RIG R G LSGGQAQ+ALAR  
 Sbjet: 417 LGNNIRFYHPSASADITRAAASAGLTVNVLPGDLEGRIGCB3GRALSGGQAQVALAR 476

Query: 481 AFLDQNRIRRVWFDEPTAHLDIETELKEIMPLASDRIVIPATHRLHMLNQMIVVME 540  
 AFL NR I++ DEPTAHLDIETE+EKE ML L D+LV ATRHLH+ MD I+V+ +  
 Sbjet: 477 AFLG-NRPILLDEPTAHLDIETEY+IKETMLDLPEDKLVIPATHRLHMLNQMDEIIVLD 535

Query: 541 KGRVAVESYQELIAKKGYLYQLKHWGG 569  
 GRVAB+G++ ELL K G +L A G  
 Sbjet: 536 GGRVAEIGTHWELLEKNGVYTKLVKAGL 564

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4981> which encodes the amino acid sequence <SEQ ID 4982>. Analysis of this protein sequence reveals the following:

40 Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

|             |                     |               |                        |
|-------------|---------------------|---------------|------------------------|
| INTEGRAL    | Likelihood = -10.61 | Transmembrane | 159 - 175 ( 154 - 190) |
| INTEGRAL    | Likelihood = -10.03 | Transmembrane | 70 - 86 ( 63 - 91)     |
| INTEGRAL    | Likelihood = -3.03  | Transmembrane | 282 - 298 ( 282 - 301) |
| 45 INTEGRAL | Likelihood = -1.44  | Transmembrane | 261 - 277 ( 260 - 278) |

----- Final Results -----

|                        |     |                                |         |
|------------------------|-----|--------------------------------|---------|
| bacterial membrane     | --- | Certainty=0.5246 (Affirmative) | < succ> |
| bacterial outside      | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| 50 bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

The protein has homology with the following sequences in the databases:

>GP:AAC22320 GB:U32749 ATP-binding transport protein (cydD)  
 [Haemophilus influenzae Rd]

55 Identities = 167/544 (30%), Positives = 279/544 (50%), Gaps = 15/544 (2%)

Query: 46 MISFYLIKTFSTFILGHATAIGRIAGLLILLNNVGVFLAIAIGK---GLQGIASQFVDS 102  
 +S+ L A F L A + LG + L L A GK O AS +  
 Sbjet: 17 VFSYILQAYFHELSLSAVILGIVLIAIAIR-----APAGKKSQVASYFASTKVKH6 70

Query: 103 LKQSFPERFIDLSQGFDAHASDADILTLASOGIDSLDTYGYGIL-SLSMRKNTCTTIMI 161  
 I+ + + + S + I+ +AS+G++ L Y+G YL L T  
 Sbjet: 71 LRSLIYRKILASHPANQVQKQSSIIQVASSGVQLEIYPCGYLPLPFLSYLLAPLTFAP 130

65 Query: 162 LVFLIYPLAGLVFLGVLPILPLSIVAMQKRSQPNMASHVSSYMDVGNLPMDDLEGILNLY 221

-1797-

L+F + A ++ L +PLIP+SI+A+ K ++ ++ YMS Y+ +G+ F+D+L+GL TL  
 Sb|ct: 131 LIPFSPKTA-IILLICVPLIPMSI IAVNKIAKLLAKYMSIYVGLSSFLNMQGLITLK 189

Query: 222 SYQATERYRQEFSGKAEQFRKATMSLLGSQLQAVGYMDAVMYLIGLGGFLAVALATGQ 281  
 YQ + +AE FRK TM +L QL +V MD + Y G + A +  
 Sb|ct: 190 IYQDDAYKAKAMDKRAHFPRKITMVKLTQMSVSLMDLLAYOGAAGILITALLQPNQAQ 249

Query: 282 LSPFNFLFFLLIATEFTFTPIREQGYMHLVMMYTMADRIIPFSLDS-VFARKMKSKYAI 340  
 LS + F+L+++EFF P+R G H+ M +D+IF+ LD+ V ++ A  
 Sb|ct: 250 LSVLGVILFILLSSSEFFIPLRLGSPFHVAMNGKASDKIPTLLDTPTVETQCSAVDFEAK 309

Query: 341 NFMQIDIQMSLAY-EKKTIVLSGVIMTLTKGQUTAIAGVSGQKSTLAQLLLKRGSAATG 399  
 N Q++I+++ +Y E+K +G+ +++ QL+ G SG GK++L LL+ A G  
 Sb|ct: 310 NNQVQIRIKILHPSYSEKPAITGLNLILPQLSVFVGKSGCKSTLAVSLMGPNQAQCG 369

Query: 400 EILFDGLSDNLSQETINQVLYVSDOSTLIARSIYDNLRLA-ANLSKKSIILWDQHKL 458  
 ILF+G ++ N+ + + Q+V VS S + ++ N+ +A + + ++I ++Q L  
 Sb|ct: 370 EILPNGQNALNIDRTSFYQKSVLSVSHSSYFVGKTLRENMTWAKIDATDQBITACIAGVNL 429

Query: 459 LSPINMLPDGLDTIVGENGLSPGQKQVICAALLSKRSLYIFDEATSSLDABENRII 518  
 F+ GLD + G LS GQ Q++ ARALL LYIFDEATS++D E+E II  
 Sb|ct: 430 AQFVR-DNGGLMDQLLSRGANLGGQIQRLALARALLNAELLYIFDEATSNIDVSEBII 488

Query: 519 DNLITRLAKTAIVIVITHMSKRLKGANQVLFNTGQFACLGKPCDLYRDQPTHYRLWDTQ 578  
 I + + +++I+H+++ A+ + L+ G+ G +L Q Y + +  
 Sb|ct: 489 LQFIQFQKQKQKTIYMSHRLANAVNADCNVLDQKLTQGTSHKELMKQYAEVPCQQ 548

Query: 579 ARLE 582  
 LE  
 Sb|ct: 549 KDLE 552

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/552 (25%), Positives = 260/552 (46%), Gaps = 12/552 (2%)

Query: 1 MLDKAVMLRSGIHKLGLLGLDVLQAIPIGQAYVLSLSITGWSQKSLSSQTVYILLF 60  
 +L + R++ LL + AL LQ + + Y + + + + G + + + LL  
 Sb|ct: 16 LLKRLRERIAPKRYLYVSAFLSWLQFVMRMISFYLIAKTFSTFILLGHATAIGRLAGLL 75

Query: 61 MVSYLGRHVIDIYINRKLDLDFSTAQSSLLRRQLDKLFDLGPVKVQKGTGNVVMALDG 120  
 +++ +G V+ + + S L++ + DL + ++T+A G  
 Sb|ct: 76 LINVVG-FVLAIGKLGQIASQFARDLKSQSFPEFIDLQGFDAHASADIITLASQ 134

Query: 121 VSLVENYLRVLNRMNINMSIIPWILAYIFYLDIERSGAILLIVFELIIPMILGZAQA 180  
 + + Y Y L+ + I+ +F+ +G+ L V PLI + ++ + Q  
 Sb|ct: 135 IDSLEYTYGYVLSLSMRKTQNCITMILVFLIPIAGLVFLGVLEPLIPLSIVAMQKRSQP 194

Query: 181 KADQYBSYQVLSNHLFDSLEGDTLKYFGLSKRYGKSIYQTSSEFRKATMSLTKIGLIS 240  
 + SY + N F+D L+G++TL + ++RY + +E FRKATMS L + +  
 Sb|ct: 195 NMSEHYSSYMDVGNLPMDDLKGLNTLYSYQATRYRQEFSGKAEQFRKATMSLLGSQLQA 254

Query: 241 TFAFDLFTTSLIAVAVFLGLRLNQLQIYLLPAITLILSPYFLPVRFDSSDYHATLDG 300  
 +D L I + L Q+ L L+++ E+F P+R+ H +  
 Sb|ct: 255 VGYMDAVMYLIGLGSQFLAVQLATGQLSFFNFLLFLLIATEFTFTPIREQGYMHLVMM 314

Query: 301 KNAFQRIQKVIKNGIKRGRLVIDWSKE----SRLDLENIAIATDQKRVEDVTLFRG 356  
 I L+ + D+ SK +++D++NI++AY++K V+ VT+  
 Sb|ct: 315 TMADRIIPFSLDSVPARK----DNKSKTAINPQIDIQINSLAYEKKTVLSGVIMTLTK 369

Query: 357 HQKVALVGVSGSKSSLINLLSGFLGPDNGSLKXVIGREVINLDQSHWHKMIYIPQTPYV 416  
 Q A+ GVSQ GK+SL LL G + DG + NL GK ++Q++Y+ +  
 Sb|ct: 370 GQUTAIAGVSGQKSTLAQLLLKRGSAATGHIILFDGLSDNLSQETINQVLYVSDQSTL 429

Query: 417 FEMSLRNITPYTPNASDEEVRAIHWGLDLSLSELPDGLTEIRIGNGARPLSGGQAQRI 476  
 S+ DN+ N S++ + I GL S + +LFDGL+T +G LS GQ ++  
 Sb|ct: 430 LNRSIYDNLRLA-ANLSKKSIILWDQHGLLSPINMLPDGLDTIVGENGLSPGQKQV 488

Query: 477 ALARAFLDQNRIMVDEPTEAHLIDTELELKERKMLPMSDRIVIFATHRLHNLQMDVI 536

-1798-

ARA L + R + +FDE T+ LD E E + + L +VI TH++ L + +  
 Sbjct: 489 ICARALLSK-RSLYIFDEATSSLDSENRRIIDNLITRIAKTAIVIVITHKMSRLKGANQV 547

Query: 537 VVMEKGRVMEVG 548

+ + G+ A +G

Sbjct: 548 LFIATGQPAICLG 559

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 1615

A DNA sequence (GBSx1710) was identified in *S.galactiae* <SEQ ID 4983> which encodes the amino acid sequence <SEQ ID 4984>. This protein is predicted to be transport ATP-binding protein cydd (cydC).

Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence  
 15 INTEGRAL Likelihood = -12.04 Transmembrane 260 - 276 ( 258 - 284)  
 INTEGRAL Likelihood = -9.34 Transmembrane 172 - 188 ( 147 - 199)  
 INTEGRAL Likelihood = -6.53 Transmembrane 150 - 166 ( 147 - 171)  
 20 INTEGRAL Likelihood = -6.05 Transmembrane 31 - 47 ( 29 - 52)  
 INTEGRAL Likelihood = -3.35 Transmembrane 68 - 84 ( 67 - 84)  
 INTEGRAL Likelihood = -1.17 Transmembrane 293 - 309 ( 292 - 310)  
 INTEGRAL Likelihood = -0.69 Transmembrane 494 - 510 ( 493 - 510)  
 ----- Final Results -----  
 25 bacterial membrane --- Certainty=0.6137 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10127> which encodes amino acid sequence <SEQ ID 10128> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15899 GB:Z99123 ABC membrane transporter (ATP-binding  
 protein) [Bacillus subtilis]  
 Identities = 262/573 (45%), Positives = 389/573 (67%), Gaps = 14/573 (2%)  
 35 Query: 16 LKTDQNIKPFKKYKYLALVIAELFLGEMTFPSASALMFNSGYLSKASLPSNILLVYVPI 75  
 +K +NI P+ KQ V+ +FLG +T FSA+ LMF SG+LISK+A+ P NILL+YVPI  
 Sbjct: 1 MKKEEWILFYIKQKARLFVLVIFLGAVTIFSAALFMPTSGFLISKAAITPENILLIYVPI 60  
 40 Query: 76 VLTRAFGIGRPVFRYIERLITSHNWVLRMTSQRLKLYHSLESNAIPMKRDFRLGDVNGLL 135  
 V R FGI R V RY+ERL H+ +L++ S +R++LY+ LE A+ ++ FR GD++G+L  
 Sbjct: 61 VAVRTFGIARSVSRVYVERLVGHILIKIVSDMRVRYNMLEPCALMLRSRFTDMLGIL 120  
 45 Query: 136 AEDINYLQNLVLRITPTTIIAMILYSFIILATGFPSSLWFALMMLLYLAIMIFLPLMSIL 195  
 +EDI +LQ+ +LVTIFP I A +LY+ +IA GFPS FA+++ LYL +++ LFP+ S+L  
 Sbjct: 121 SEDIBHLQDAFLKTI FPAISALLVAVSVIALGFPFSDWPFALLALYLFPVIVLFPVVSLL 180  
 Query: 196 ANGARQTREREKLNHLYTDLTDNLVLGISDWI FSGRGQRYVALHERSESEIMVQKIRSF 255  
 A+ + K +N LY+ LVD V+G+SDW+FS R ++ +E+ E + +++K + F  
 50 Sbjct: 181 VTRAKNAKLGKGNVLYSRITDAVNGVSDWPFSGRRHAFIDAYEKEERDWFELERKKQRF 240  
 Query: 256 DNRRALIVELVFGFLAILVLIWAAGNQPIGHGGEA--NWIAAFVLTVFPLSRAPAGLSAA 313  
 R + + L +L++ W + Q GE IAAFLV VFPL+EAF LS A  
 Sbjct: 241 TRWRDFAAQCLVAGLILMLFWTAGQ---QADGLAKTMIIAFVLVVFPLTEFLPLSDA 297  
 55 Query: 314 AQETNKYSDSIHRLN-----ELSETYPETTONQLPNKPYDFSVKNLSQPYQKQKRWLH 367  
 E Y DSI R+N E S+T E L + + +++F Y + VLIH  
 Sbjct: 298 LGEVTPGYQDSIRRNWVAQPEASQT--ESGDQILDLQDVTLAFRDVTPSYDINSQ-VLIH 354